MIRCEA BIJI

ELENA BIJI

STATISTICĂ TEORETICĂ

EDITURA DIDACTICĂ ŞI PEDAGOGICĂ, BUCUREŞTI — 1979

Scanned with OKEN Scanner

ACADEMIA DE STUDII ECONOMICE — BUCUREŞTI

Prof. univ. **MIRCEA BIJI**
Membru corespondent al
Academiei R. S. România

Prof. univ. **ELENA BIJI**
Doctor în economie

STATISTICĂ TEORETICĂ

Scanned with OKEN Scanner

CUVÎNT ÎNAINTE

Alcătuit pe baza programei analitice a cursului de Statistică teoretică de la Facultatea de Planificare și cibernetică economică a Academiei de studii economice din București, manualul cuprinde problemele principale ale metodologiei statisticii social-economice, selectate în funcție de nevoile construcției socialismului, în prezent și în perspectivă, în țara noastră. Autorii s-au străduit să folosească experiența lor didactică îndelungată pentru selectarea problemelor esențiale ale metodei statistice aplicate în studiul fenomenelor social-economice, urmărind treptele cunoașterii statistice, stăruind asupra procedeelor moderne pe care le impune practica construcției economice în condițiile revoluției tehnico-științifice. Această tratare selectivă și nu exhaustivă a metodologiei statistice nu împiedică parcurgerea întregului proces de culegere, prelucrare și analiză a informațiilor statistice, necesar elaborării sistemului informațional al economiei naționale și analizei complexe a fenomenelor economice și sociale.

Principiile generale și problemele metodologice ale culegerii și prelucrării primare a informațiilor statistice au fost sintetizate în manual, într-un spațiu relativ restrîns. Prelucrarea indicatorilor derivați și procedeele de analiză statistică constituie partea principală a manualului.

Orientarea pentru partea practică, aplicativă este ilustrată prin modele de calcul și analiză statistică pe baza datelor concrete ale dezvoltării societății noastre socialiste. Criteriile de selectare a informațiilor statistice concrete au fost determinate de cerințele metodei aplicate și se referă fie la cicluri cincinale, fie la ultimele date publicate de organele de specialitate sau care se găsesc în documentele de partid și de stat. Prioritatea acordată problemelor metodologice, potrivit profilului manualului, nu a împiedicat reliefarea aspectelor de conținut economic și social cuprinse în datele statistice analizate.

Manualul de Statistică teoretică este destinat, în primul rînd, studenților din învățămîntul economic superior. El poate fi la fel de util studenților din oricare specialitate care are contingență cu fenomene aleatoare. Ținînd seama de vastul sistem organizat în țara noastră pentru perfecționarea permanentă a pregătirii lucrătorilor din toate domeniile de activitate, autorii s-au străduit să structureze astfel manualul încît să fie folositor nemijlocit tuturor acelora care lucrează și se perfecționează continuu în variatele domenii ale activității sociale și economice pe diferite trepte ierarhice.

AUTORII.

Scanned with OKEN Scanner

CUPRINS

Cuvînt înainte 3

Cap. I .NOȚIUNI INTRODUCTIVE ÎN STATISTICA SOCIAL-ECONOMICĂ 11

1.1. Scurtă prezentare istorică a dezvoltării statisticii 12
1.2. Sfera de cuprindere a statisticii social-economice 18
1.3. Particularități ale obiectului de studiu al statisticii social-economice 20
1.4. Metoda statisticii 23
1.5. Principalele noțiuni folosite în statistica social-economică 25
1.6. Organizarea cercetării statistice 27
1.7. Funcțiile statisticii în economia socialistă 30

Cap. II. OBSERVAREA STATISTICĂ 31

2.1. Organizarea unei observări statistice 31
2.2. Clasificarea observărilor statistice 35
2.3. Tipuri de observări statistice 36
2.3.1. Sistemul dărilor de seamă 36
2.3.2. Recensămîntul 38
2.3.3. Observarea selectivă 39
2.3.4. Observarea părții principale 39
2.3.5. Anchetele statistice 39
2.3.6. Monografia 40
2.4. Erorile observării statistice 40
2.5. Controlul datelor statistice 41

Cap. III. PRELUCRAREA DATELOR STATISTICE 42

3.1. Planul prelucrării statistice 43
3.2. Tehnici de prelucrare 44
3.3. Centralizarea datelor statistice 45
3.3.1. Centralizarea simplă 45
3.3.2. Centralizarea pe grupe 46
3.4. Metoda grupării — metodă de bază în prelucrarea datelor statistice 46
3.4.1. Alegerea și folosirea caracteristicilor de grupare 48
3.4.2. Felurile grupărilor statistice 48
3.4.3. Alegerea numărului de grupe și stabilirea mărimii intervalului de grupare pentru caracteristicile exprimate numeric 50
3.4.4. Funcțiile grupării statistice 53
3.5. Folosirea indicatorilor statistici absoluți și derivați în prezentarea și analiza rezultatelor centralizării și grupării 56

Cap. IV. PREZENTAREA DATELOR STATISTICE 58

4.1. Tabele statistice 58
4.1.1. Elementele tabelului statistic 58
4.1.2. Felurile tabelelor statistice 59
4.1.3. Reguli pentru întocmirea și interpretarea tabelelor statistice . . 63
4..2 Seriile statistice 64

Scanned with OKEN Scanner

4.2.1. Feluri de serii statistice 64
4.2.2. Elemente de analiză a seriilor statistice 70
4.3. Prezentarea grafică a datelor statistice 72
4.3.1. Elementele de bază ale unui grafic 73
4.3.2. Principalele tipuri de grafice utilizate în statistica social-economică 74
Cap. V. MĂRIMI RELATIVE 89
5.1. Noțiunea de mărime relativă în statistica social-economică 89
5.2. Felurile mărimilor relative 91
Cap. VI. MĂRIMILE MEDII, INDICATORII VARIAȚIEI ȘI ASIMETRIEI 100
6.1. Mărimile medii în statistica social-economică. Condiții de aplicare 100
6.2. Felurile mediilor utilizate în statistica social-economică 102
6.2.1. Media aritmetică simplă și ponderată 103
6.2.1.1. Proprietățile mediei aritmetice 105
6.2.1.2. Calculul mediei într-o serie de distribuție de frecvențe pe intervale de grupare 111
6.2.1.3. Calculul simplificat al mediei aritmetice ponderate în seriile de intervale 113
6.2.1.4. Media caracteristicii alternative 116
6.2.2. Media armonică simplă și ponderată 118
6.2.3. Media pătratică simplă și ponderată 121
6.2.4. Media geometrică simplă și ponderată 122
6.2.5. Calculul mediei unor indicatori derivați, exprimați prin mărimi relative parțiale sau mărimi medii parțiale 125
6.2.6. Valori medii de poziție sau de structură 128
6.2.6.1. Mediana 128
6.2.6.2. Modulul 131
6.2.6.3. Cuartile, decile 132
6.3. Indicatorii variației și asimetriei 134
6.3.1. Indicatorii simpli ai variației 136
6.3.2. Indicatorii sintetici ai variației 137
6.3.3. Calculul simplificat al dispersiei și abaterii medii pătratice. Proprietățile dispersiei 143
6.3.4. Indicatorii variației într-o colectivitate împărțită în grupe. Regula adunării dispersiilor 148
6.3.5. Calculul indicatorilor variației unei caracteristici alternative . . . 154
6.3.6. Variație intercuartilică și interdecilică 156
6.4. Asimetria 158
6.4.1. Interpretarea asimetriei pe baza proprietăților distribuției normale 159
6.4.2. Indicatorii asimetriei 163
Cap. VII. CERCETAREA SELECTIVĂ 167
7.1. Noțiuni introductive 167
7.2. Procedee de selecție 173
7.2.1. Selecții aleatoare 173
7.2.1.1. Procedeul tragerii la sorți 173
7.2.1.2. Procedeul tabelului cu numere întîmplătoare 174
7.2.1.3. Selecția mecanică 175

Scanned with OKEN Scanner

7.2.2. Selecții dirijate 175
7.3. Erorile cercetării selective 177
7.3.1. Tipuri de erori 177
7.3.1.1. Eroarea efectivă de reprezentativitate 179
7.3.1.2. Eroarea medie probabilă de reprezentativitate 182
7.3.1.3 Eroarea limită de reprezentativitate 191
7.4. Felurile selecțiilor utilizate frecvent în practica statistică. Calculul erorilor de reprezentativitate pe feluri de selecție 194
7.4.1. Selecția întîmplătoare simplă 194
7.4.2. Selecția mecanică 195
7.4.3. Selecția tipică (stratificată) 196
7.4.4. Selecția de serii 199
7.4.5. Alte feluri de selecție folosite în practica statistică 201
7.5. Determinarea volumului de selecție 203
7.6. Selecția de volum redus 210
7.7. Extinderea rezultatelor selecției asupra colectivității generale . . 212
Cap. VIII. ELEMENTE DE ANALIZĂ DISPERSIONALĂ 217
8.1. Indicatorii utilizați în analiza dispersională 219
8.2. Aplicații ale analizei dispersionale în cazul unei grupări simple . . 223
Cap. IX. ANALIZA LEGĂTURILOR DINTRE FENOMENELE SOCIAL-ECONOMICE 228
9.1. Tipuri de legături între fenomenele social-economice 229
9.2. Metode și procedee de verificare a legăturilor statistice 232
9.2.1. Procedeul seriilor statistice interdependente 233
9.2.2. Metoda grupărilor 234
9.2.3. Metoda grafică 237
9.2.4. Procedeul tabelului de corelație 239
9.2.5. Metoda balanțelor 243
9.3. Metode analitice de măsurare și analiză a legăturilor statistice . . . 244
9.3.1. Corelația liniară simplă 244
9.3.1.1. Noțiunea și calculul ecuațiilor de regresie 245
9.3.1.2. Raportul de corelație 250
9.3.1.3. Coeficientul de corelație 259
9.3.1.4. Relații între coeficientul de corelație liniară și parametrii ecuației de regresie 261
9.3.2. Corelația simplă curbilinie 264
9.3.2.1. Determinarea ecuațiilor de regresie 264
9.3.2.2. Raportul de corelație 268
9.3.3. Corelația multiplă în cazul distribuțiilor multidimensionale de formă liniară 269
9.3.3.1. Determinarea ecuațiilor de regresie 270
9.3.3.2. Raportul de corelație 272
9.3.3.3. Corelația parțială 278
9.3.4. Metode neparametrice de măsurare a legăturilor dintre fenomenele social-economice 280
9.3.4.1. Tabelul de asociere și coeficientul de asociere 280
9.3.4.2. Coeficienții de corelație a rangurilor 282
9.3.5. Analiza raporturilor de corelație cu decalaj 286

Scanned with OKEN Scanner

Cap. X. ANALIZA SERIILOR CRONOLOGICE 289
10.1. Noţiuni. Particularităţi 299
10.2. Felurile seriilor cronologice 291
10.3. Prelucrarea seriilor cronologice de intervale 293
10.3.1. Indicatorii absoluţi care caracterizează seriile cronologice . . . 294
10.3.2. Indicatorii relativi ai seriilor cronologice 297
10.3.3. Indicatorii medii ai unei serii cronologice de intervale 303
10.4. Prelucrarea seriilor cronologice de momente 308
10.4.1. Media cronologică simplă 309
10.4.2. Media cronologică ponderată 311
10.5. Ajustarea seriilor cronologice. 313
10.5.1. Ajustarea pe baza mediilor mobile 315
10.5.2. Ajustarea pe baza metodei grafice 317
10.5.3. Ajustarea pe baza sporului mediu 318
10.5.4. Ajustarea pe baza indicelui mediu de creştere 321
10.5.5. Ajustarea prin metoda celor mai mici pătrate 324
10.5.6. Criterii de alegere a procedeelor de ajustare 333
10.6. Măsurarea oscilaţiilor sezoniere 336
10.6.1. Calculul sezonalităţii folosind metoda mediilor aritmetice . . . 337
10.6.2. Calculul sezonalităţii folosind metoda mediilor mobile 339
10.6.3. Calculul sezonalităţii pe baza datelor ajustate prin metoda celor mai mici pătrate 341
10.7. Folosirea datelor statistice pentru fundamentarea calculelor de prognoză 344
10.7.1. Extrapolarea datelor statistice 346
10.7.1.1. Extrapolarea prin procedee mecanice 347
10.7.1.2. Extrapolarea pe baza metodelor analitice de calcul 349
10.7.2. Interpolarea seriilor cronologice 352
10.7.3. Calcule de perspectivă pe plan teritorial 355
Cap. XI. INDICII 361
11.1. Clasificarea indicilor. 362
11.1.1. Indicii dinamicii, teritoriali şi ai planului 362
11.1.2. Indicii individuali şi de grup 363
11.1.3. Indicii agregaţi, sub formă de medie şi sub formă de raport de medii 364
11.1.4. Indicii cu structură variabilă, structură fixă şi ai variaţiei structurii 366
11.1.5. Indicii cu bază fixă şi cu bază variabilă 370
11.1.6. Indicii cu ponderi constante şi cu ponderi variabile 370
11.2. Sisteme de ponderare folosite la construirea indicilor. Relaţia lui Bortkiewicz 373
11.3. Descompunerea pe factori a variabilei unui fenomen complex folosind metoda indicilor 378
11.4. Aplicaţii la construirea unor indici economici 383
11.4.1. Indicii valorii, volumului fizic şi ai preţurilor producţiei şi circulaţiei mărfurilor 383
11.4.2. Indicii productivităţii muncii 394
11.5. Particularităţi în construirea unor indici 405
11.5.1. Indicii sarcinii de plan şi ai îndeplinirii planului 406
11.5.2. Indicii teritoriali 409

Scanned with OKEN Scanner

Cap. XII. PROBLEME METODOLOGICE ALE INTERPRETĂRII DATELOR STATISTICE 413
12.1. Bazele teoretice şi sarcinile interpretării datelor statistice 413
12.2. Sursele de date utilizate în analiza statistică 416
12.3. Metode şi procedee de interpretare a datelor statistice 416
12.4. Expunerea rezultatelor interpretării datelor statistice 421
Anexe 422
Bibliografie 429

Scanned with OKEN Scanner

CAPITOLUL I

NOȚIUNI INTRODUCTIVE ÎN STATISTICA SOCIAL-ECONOMICĂ

Complexitatea proceselor social-economice, marea varietate a formelor concrete prin care se manifestă fenomenele pe care acestea le generează, impun folosirea combinată a celor mai variate metode de investigație științifică necesare cunoașterii legilor obiective de dezvoltare a societății și a particularităților sub care ele se manifestă. Studiul acestor fenomene de tip colectiv — variabile în timp și spațiu — ale căror proprietăți nu pot fi cunoscute decît la nivelul întregului, al ansamblului, au impus statistica ca o disciplină științifică independentă.

Folosirea statisticii ca „instrument de cunoaștere a vieții social-economice" — expresia o găsim la Lenin [1] — nu este de dată recentă. Statistica a fost folosită pentru rezolvarea unor nevoi practice ale vieții sociale din cele mai vechi timpuri și pînă astăzi, o regăsim atît în primele forme de evidență a populației și bunurilor materiale, cît și în rezolvarea, în condițiile actuale ale folosirii echipamentelor moderne de calcul, a celor mai variate și complexe probleme de conducere.

Izvorîtă din necesitățile practice ale societății, statistica s-a dezvoltat continuu, participînd din ce în ce mai mult la procesul de cunoaștere și transformare a naturii și societății.

În economia socialistă, statistica capătă funcții noi, devenind pivotul sistemului informațional, iar prin concepte, metode, tehnici și operanzi specifici, contribuie direct sau prin intermediul altor științe la procesul de conducere planificată a economiei naționale.

Se poate spune, deci, că statistica are un rol important în procesul de cunoaștere a stărilor din trecut și prezent și, mai ales, în studierea tendințelor obiective ale dezvoltării sociale și economice, fiind folosită la elaborarea prognozelor din cele mai variate domenii ale economiei naționale. În raportul prezentat la cel de-al XI-lea Congres al P.C.R., tovarășul Nicolae Ceaușescu, folosind un bogat material statistic, spunea: „Ca urmare a aplicării neabătute a politicii de industrializare socialistă — baza dezvoltării întregii economii, a progresului general al țării și a asigurării unei reale independențe naționale — producția industrială a crescut în această perioadă de circa 30 de ori. Estimată în prețuri comparabile, producția industrială a fost în 1938 de circa 18 miliarde lei, în 1950 de 26 miliarde, în 1955 de 53 miliarde, în 1960 de 89 miliarde, în 1965 de 170 miliarde, în 1970 de peste 300 miliarde, iar în 1975 — pe baza realizărilor din primii patru ani ai cincinalului și a prevederilor de plan pe anul viitor — va fi de aproape 580 miliarde lei" [2]. Această prevedere a fost îndeplinită și depășită obținîndu-se 586,9 miliarde lei, ceea ce înseamnă un

---

[1] V. I. Lenin, *Opere*, vol. 17. E.S.P.L.P., București, 1957, p. 435.

[2] Nicolae Ceaușescu, *Raport la cel de-al XI-lea Congres al Partidului Comunist Român*, Editura politică, București, 1974, p. 7—8.

grad de îndeplinire a planului de 101,2% [1]. Pe aceeaşi linie se înscriu şi succesele obţinute în primii trei ani ai cincinalului 1976—1980, succese care au confirmat pe deplin realismul hotărîrilor şi orientărilor adoptate de Congresul al XI-lea, justeţea prevederilor Programului partidului. Industria s-a dezvoltat într-un ritm mediu anual de peste 11 la sută faţă de 10 la sută cît se prevăzuse iniţial, asigurîndu-se astfel o producţie suplimentară pe cei trei ani de circa 60 miliarde lei.

Pentru a înţelege mai bine rolul statisticii în procesul de cunoaştere a vieţii sociale este necesar să se facă o succintă trecere în revistă a apariţiei şi dezvoltării sale istorice.

1.1. SCURTĂ PREZENTARE ISTORICĂ A DEZVOLTĂRII STATISTICII

Statistica, în sensul larg de evidenţă a fenomenelor şi proceselor sociale şi economice, a apărut cu mult înaintea utilizării termenului. Termenul de „statistică" apare în secolul al XVIII-lea şi îşi are originea în latinescul „status" cu sensul de situaţie, stare, sau în cuvîntul italian „stato" care are aceeaşi rădăcină şi are înţelesul de stat; deci, iniţial, noţiunea nu îmbracă decît în linii foarte largi ceea ce se înţelege azi prin statistică, atît ca disciplină ştiinţifică, cît şi ca activitate practică.

Din cercetările istorice rezultă că primele forme de evidenţă au apărut la cele mai vechi colectivităţi omeneşti, fiind folosite mai ales în perioada de descompunere a comunei primitive şi a apariţiei statului sclavagist.

Apariţia proprietăţii private şi a statului sclavagist a impus o cunoaştere amănunţită a numărului de sclavi, a numărului de persoane capabile să poarte arme, precum şi a mărimii averilor şi a numărului persoanelor impozabile. Alături de evidenţa statului sclavagist apar şi unele evidenţe ale stăpînilor de sclavi.

Astfel, în antichitate se întîlnesc forme de evidenţă ce pot fi asimilate înţelesului modern de recensăminte statistice, în special în China, Egipt, Grecia şi în Imperiul Roman. Din documente rezultă că chinezii dispuneau încă din mileniul al IV-lea şi al III-lea î.e.n. de date statistice cu privire la numărul populaţiei, structura terenurilor şi că utilizau diferite tabele statistice referitor la unele aspecte ale activităţilor agricole. În vechiul Egipt s-au găsit forme de evidenţă care aparţin perioadei de construire a piramidelor, izvorîte din nevoia de a se ţine evidenţa sclavilor necesari pentru munca la aceste construcţii uriaşe, a cantităţii de hrană necesară, cît şi a cantităţii de materiale de construcţie utilizate. Grecii şi romanii foloseau curent recensămintele ca formă de evidenţă în scopuri fiscale şi militare. Ei organizau recensăminte pentru cunoaşterea repartiţiei terenurilor şi a condiţiilor de locuit în vederea fixării impozitelor, precum şi a cunoaşterii numărului de persoane în stare de a purta arme. După Tacit, împăratul Augustus a fost primul care a făcut un recensămînt culegînd date cu privire la averile particulare şi la veniturile publice, precum şi date cu privire la numărul soldaţilor şi corăbiilor.

[1] *Anuarul statistic al R.S. România*, 1977, D.C.S., p. 84.

Scanned with OKEN Scanner

Recensămîntul (censul) populaţiei la romani avea un caracter periodic, efectuîndu-se din 5 în 5 ani, iar mai tîrziu din 10 în 10 ani.

Din documentele descoperite rezultă că lucrări asemănătoare de evidenţă erau folosite şi pe teritoriul de azi al ţării noastre; administraţia romană în Dacia a introdus şi aici lucrări asemănătoare de evidenţă a populaţiei, producţiei şi consumului, organizînd servicii speciale de evidenţă (tabularium).

Această lungă perioadă istorică, care corespunde primei etape a apariţiei şi dezvoltării statisticii, se poate caracteriza prin faptul că se cunosc numai forme generale de evidenţă, în majoritatea lor cu caracter statistic (recensăminte, înregistrări demografice din vechile temple cu privire la numărul naşterilor şi al deceselor etc.).

Odată cu trecerea la relaţiile feudale de producţie, pe măsura dezmembrării vechilor imperii din antichitate şi a apariţiei principatelor independente, s-a simţit nevoia organizării unor evidenţe mai cuprinzătoare pentru fiecare feudal în parte, cu privire la mărimea bunurilor posedate, a rezultatului producţiei şi a evidenţei proprietăţii incomplete asupra iobagilor de pe feude. În orînduirea feudală, în prima sa etapă, forma specifică de evidenţă o constituie inventarul pentru marii latifundiari, alături de care se întîlnesc formele de evidenţă ale statului feudal sub formă de catagrafii, cu privire la structura populaţiei şi a proprietăţilor, întocmite în special în scopuri fiscale.

Forme asemănătoare de evidenţă se utilizează în această perioadă şi în ţara noastră, atît în Muntenia şi Moldova, cît şi în Transilvania. În Muntenia şi Moldova se cunosc unele lucrări de evidenţă pe lîngă visteria domnitorului, sub formă de catastihe săteşti, iar în timpul domniei lui Constantin Mavrocordat lucrări cu caracter statistic denumite catagrafii.

Pentru Transilvania există o documentaţie bogată privitoare la această perioadă, păstrată în arhivele locale sau în arhivele de la Viena, fie sub formă de „urbarii", fie mai tîrziu, de „conscripţii" cu caracter fiscal. Prin „urbarii" se înţelegea descrierea unei localităţi împrejmuite de un hotar, prezentată sub formă de tabel sau de înşiruire de date cu care se consemnau în special multiplele obligaţii ale iobagilor faţă de feudal. În „conscripţiile" fiscale se înscriau obligaţiile contribuabililor către stat şi comitat, adică impozitele. Aceste forme de evidenţă cuprindeau date cu privire la forţa de muncă — iobagii de pe moşii —, iar mai tîrziu, pe măsura apariţiei formelor muncii salariate, date cu privire la cei ce lucrau în manufacturi şi comerţ.

Apariţia atelierelor, breslelor, oraşelor, a comerţului interior şi exterior a determinat transformări importante în conţinutul şi caracterul evidenţei. De la evidenţa care avea un caracter limitat la fiecare feudă în parte odată cu dezvoltarea forţelor de producţie se fac încercări de organizare a unei evidenţe la nivelul întregului stat. În acelaşi timp, dezvoltarea comerţului imprimă un caracter nou evidenţei, sub forma evidenţei contabile. Apar primele cursuri de evidenţă ale lui S. Catrugli (1458) şi Luca Pacciolo (1494). Se dezvoltă, de asemenea, evidenţa cu caracter demografic ţinută de preoţi, care aveau sarcina de a culege informaţii despre căsătorii, naşteri, decese.

Scanned with OKEN Scanner

În aceste condiţii are loc o primă diferenţiere a formelor generale de evidenţă: contabilitatea, cu tendinţa de a ţine gestiunea în interesul proprietarilor particulari (negustori, moşieri etc.) şi statistica, cu tendinţa de a servi prin culegerea de date şi informaţii conducerea de stat.

Aceasta reprezintă a doua fază a dezvoltării statisticii şi se caracterizează prin apariţia statisticii descriptive şi trecerea la aritmetica politică.

În secolele al XV-lea şi al XVI-lea abundă lucrări în care se folosesc numeroase date statistice pentru descrierea amănunţită a situaţiei social-economice a diferitelor state. Astfel de lucrări se găsesc aproape în toate statele europene şi sînt cunoscute sub denumirea generică de statistică descriptivă.

Printre numeroasele lucrări de statistică descriptivă se poate aminti cartea italianului F. Sansovino despre guvernare şi administraţie, care are în vedere descrierea statului cu bogăţiile sale, a populaţiei, a industriei, comerţului etc.; în Franţa, lucrarea lui J. Bodin „Republica" (1576) care, printre altele, pune problema folosirii curente a înregistrărilor numerice şi a organizării periodice a recensămintelor populaţiei; în Germania, o lucrare asemănătoare aparţinînd lui S. Münster, intitulată „Cosmografia"; în Rusia, lucrarea lui Kirilov, intitulată „Situaţia înfloritoare a statului rus", iar la noi, cunoscuta lucrare a lui Dimitrie Cantemir „Descriptio Moldaviae".

Tot în această perioadă se întîlnesc şi încercări de teoretizare a practicii statistice, cum sînt cele ale profesorului Conring, care a întocmit, dealtfel, şi primul curs de statistică intitulat „Notitia rerum publicarum" (1660).

Reluînd ideile lui Conring un secol mai tîrziu, Achenwald defineşte statistica ca o ştiinţă descriptivă folosită pentru prezentarea particularităţilor unui stat, introducînd pentru prima oară termenul de statistică (Statistik).

Apariţia şi dezvoltarea relaţiilor de producţie capitaliste în economie au avut ca rezultat modificări radicale în suprastructura societăţii. Burghezia în ascensiune, interesată în cunoaşterea cantitativă a fenomenelor sociale şi economice, a folosit statistica ca instrument în lupta sa împotriva vechilor forme ale feudalismului, imprimînd astfel statisticii un caracter progresist.

Dezvoltarea modului de producţie capitalist a impus o nouă organizare a formelor de evidenţă şi statistică, care trebuiau să caracterizeze numeric forţa de muncă, mărimea duratei zilei de muncă, suma salariilor plătite muncitorilor, mărimea şi structura costurilor de producţie, calculul amortizării fondurilor fixe şi alte aspecte legate de noua formă de organizare a producţiei.

În această perioadă apare un nou mod de a concepe statistica, datorită şcolii „aritmeticii politice" reprezentată prin W. Petty (1623—1687), autorul cunoscutei lucrări „Aritmetica politică" şi astronomului C. Halley (1662—1742).

W. Petty, în lucrarea „Aritmetica politică", apărută în 1690, a pus bazele statisticii ca ştiinţă, folosind expresia numerică în studiul fenomenelor sociale. Karl Marx l-a apreciat în mod deosebit pe W. Petty, spunînd despre el: „Metoda sa nu este cea tradiţională. În loc să lege într-un singur tot un şir de cuvinte, la comparativ şi superlativ şi să adauge la aceasta

Scanned with OKEN Scanner

argumente speculative, el a început să vorbească prin intermediul numerelor, al greutăţilor şi al măsurilor, să folosească în exclusivitate argumente luate din experienţa senzorială şi să examineze cauzele care au temei vizibil în natură" [1].

W. Petty încearcă să descopere regularităţile fenomenelor sociale şi foloseşte pentru prima dată noţiunea de mărime medie (mărimea valorii mărfurilor este determinată de munca medie şi nu de munca cheltuită în fiecare caz individual).

Pentru această perioadă, în Anglia, este de reţinut şi activitatea lui J. Graunt, care a folosit statistica pentru demonstrarea unor regularităţi în mişcarea naturală a populaţiei (proporţia între băieţi şi fete la naştere —în medie la 100 fete se nasc 105 băieţi — mortalitatea mai ridicată la bărbaţi decît la femei).

Apariţia şcolii „aritmeticii politice" în Anglia reprezintă o etapă progresistă, raportată la condiţiile istorice în care ea a apărut şi s-a dezvoltat. Şcoala aritmeticii politice se caracterizează prin depăşirea formei descriptive şi trecerea la o analiză a unui număr mare de cazuri individuale înregistrate, în vederea evidenţierii anumitor regularităţi în producerea fenomenelor sociale. Încercînd să cerceteze anumite legi pe baza datelor statistice, reprezentanţii ei formulează unele previziuni, dar neluînd în considerare condiţiile concrete în care aceste legi se manifestă, consideră societatea capitalistă ca fiind imuabilă.

Deşi apariţia acestei şcoli marchează o etapă progresistă în dezvoltarea istorică a statisticii, în special prin introducerea unor noţiuni şi categorii statistice moderne, totuşi ea nu a reuşit să contureze obiectul şi metoda de cercetare a statisticii. Caracterul limitat al şcolii aritmeticii politice constă în faptul că reprezentanţii ei nu au înţeles legitatea istorică a fenomenelor sociale, ca expresie a legăturii reciproce dintre aceste fenomene şi condiţiile concrete de timp şi spaţiu în care ele se manifestă.

Pe aceeaşi linie, a lărgirii folosirii instrumentelor matematice în statistică, se înscriu lucrările lui A. Quételet (1796—1874), considerat de istoria burgheză ca fondator al statisticii moderne. El a folosit în cercetările sale, în domeniul demografiei şi criminalităţii, calculul probabilităţii şi legea numerelor mari, fără a ţine însă seama de particularităţile specifice ale fenomenelor social-economice, de variabilitatea lor în timp şi spaţiu. Tratînd fenomenele aistoric şi nediferenţiate pe tipuri calitative existente obiectiv în cadrul societăţii, concluziile la care ajunge nu au putut fi decît eronate.

K. Marx şi F. Engels au criticat în operele lor concepţia lui Quételet, demonstrînd, pe baza studiilor făcute, că regularităţile stabilite de statistică nu sînt veşnice, că ele se schimbă în funcţie de schimbările social-economice. Aceasta înseamnă că, în cercetarea statistică, modul de interpretare a datelor poate transforma statistica într-un instrument de cunoaştere a realităţii obiective sau, din contră, într-un instrument de apologiere a unei societăţi, în funcţie de concepţia despre lume pe care statisticianul o are.

În perioada de ascensiune a capitalismului se accentuează procesul de diferenţiere a evidenţei, în special a evidenţei contabile şi statistice.

[1] K. Marx, *Contribuţii la critica economiei politice*, E.S.P.L.P., 1954, p. 48.

Scanned with OKEN Scanner

La mijlocul secolului al XIX-lea apar primele birouri oficiale de statistică, care se transformă treptat, pe măsura creşterii atribuţiilor lor, în institute centrale de statistică [1]. Ele se ocupă cu înregistrarea curentă a mişcării naturale a populaţiei, execută recensăminte care se referă atît la populaţie, cît şi la întreaga activitate economică. Treptat, statistica ajunge să cuprindă o mare parte din ramurile economiei.

Lărgirea sferei de cuprindere a fenomenelor cercetate sub raport statistic s-a reflectat şi pe plan teoretic şi metodologic. Din ce în ce mai mult, statistica nu mai este înţeleasă numai ca o tehnică de obţinere a unor date individuale numeroase şi de prelucrare a acestora, ci mai ales ca o metodă ştiinţifică cu largă aplicabilitate în cercetarea concretă. Statistica nu se mai poate limita numai la o descriere cantitativă a unor fenomene complexe, ci ea trebuie să le analizeze şi să le interpreteze prin metodele sale specifice. Apariţia în matematică a calculului probabilităţilor a dat statisticii un nou avînt, atît în ceea ce priveşte fundamentarea sa teoretică, cît şi în stabilirea metodelor proprii de cercetare şi a permis extinderea cîmpului studiilor sale în toate domeniile în care întîmplarea este considerată a juca un rol deosebit. Utilizarea calculului probabilităţilor în cercetarea statistică concretă a dat metodei statistice tocmai instrumentul necesar pentru a putea trece la formularea unor ipoteze de lucru, care să realizeze legătura dintre cercetarea empirică şi concluzia statistică.

Trecerea de la culegerea de informaţii numerice la interpretarea probabilistică a fenomenelor s-a realizat într-un proces destul de îndelungat şi s-a concretizat în lărgirea permanentă a ariei de aplicabilitate a metodei statistice şi în creşterea funcţiei sale de cunoaştere.

Începînd cu cea de-a doua jumătate a secolului trecut, statistica cunoaşte o largă dezvoltare. Apar variate curente, uneori contradictorii, în special în ceea ce priveşte modul de interpretare a rezultatelor cercetării statistice.

În literatura de specialitate continuă şi astăzi să se întîlnească moduri diferite de a concepe şi defini statistica. Toate acestea însă nu au făcut decît să stimuleze elaborarea şi perfecţionarea metodologiei de cunoaştere statistică a fenomenelor, de lărgire continuă a ariei sale de aplicabilitate, de cuprindere a ei în cadrul ştiinţelor ce abordează sistemic fenomenele şi procesele pe care le studiază.

De aceea, pentru a înţelege cît mai corect conţinutul şi rolul statisticii în cunoaşterea fenomenelor sociale şi economice este necesar ca în evoluţia istorică a statisticii moderne să fie scoase în evidenţă două momente principale — prezentate în succesiunea lor şi nu a importanţei —, aplicarea statisticii matematice, în primul rînd, a calculului probabilităţilor şi fundamentarea procesului său de cunoaştere pe materialismul dialectic.

Utilizînd ca instrumente de cercetare metodele matematice şi în special calculul probabilităţilor, statistica social-economică a putut trece pragul de la descrierea cantitativă şi formularea unor regularităţi în pro-

[1] La noi în ţară, la 28 aprilie 1859, Al. I. Cuza aprobă înfiinţarea unui birou de statistică în Ţara Românească, numind şef pe Dionisie Pop Marţian, iar la 1 iunie 1859, Direcţia de Statistică din Moldova sub conducerea lui Ion Ionescu de la Brad. La 4 august 1862 cele două organe de statistică sînt contopite sub denumirea de Oficiul Statistic pentru Principatele Unite sub conducerea lui Dionisie Pop Marţian.

Scanned with OKEN Scanner

ducerea fenomenelor sociale, la analiza lor, la cunoaşterea legilor care le determină şi, pe această bază, la efectuarea de previziuni ştiinţifice.

Efectele introducerii calculului probabilităţilor în cercetarea statistică se concretizează în :

— emiterea unor ipoteze statistice ştiinţific fundamentate, în care problema centrală nu este determinarea unei singure valori certe, ci a unui interval de valori în care trebuie să se situeze mărimea evenimentului studiat, a cărui siguranţă de producere este garantată de probabilitatea sa de apariţie :

— posibilitatea de a trece la descoperirea legilor statistice prin stabilirea unor legături stabile între frecvenţele şi probabilităţile de apariţie a fenomenelor, manifestate în tendinţa de concentrare a nivelurilor individuale către o valoare centrală tipică, ce nu poate fi interpretată decît ca rezultat al acţiunii legii numerelor mari ;

— posibilitatea de previziune a modului de desfăşurare a evenimentului cercetat în etapa care urmează, prin extrapolarea datelor trecutului şi prezentului asupra viitorului, în limitele admise de teoria estimării folosită în calculele de previziune.

În ceea ce priveşte cel de-al doilea moment hotărîtor pentru dezvoltarea statisticii — fundamentarea ei pe principiile dialecticii materialiste —, aceasta a fost de natură să modifice însuşi conţinutul statisticii, a permis conturarea obiectului său de studiu, precizarea limitelor de aplicabilitate a metodei sale, precum şi corelaţia cu celelalte ştiinţe.

Fundamentarea statisticii pe principiile dialecticii materialiste a permis trecerea de la analiza cantitativă a fenomenelor la o analiză calitativă, multilaterală, bazată pe cunoaşterea legilor obiective de dezvoltare a societăţii şi a condiţiilor concrete istorice în care ele se manifestă. În acelaşi timp, interpretînd materialist-dialectic societatea, s-a putut contura cu claritate rolul statisticii în realizarea procesului de cunoaştere a societăţii. Operele clasicilor marxism-leninismului cuprind, pe lîngă ilustrarea bogată cu date statistice a tezelor care au revoluţionat ştiinţa, indicaţii preţioase asupra rolului statisticii în cunoaşterea şi transformarea revoluţionară a societăţii.

Încă de la începuturile organizării sale, clasa muncitoare a acordat o atenţie deosebită organizării unei statistici ştiinţifice bazată pe date reale, pe care să le poată utiliza în lupta sa împotriva burgheziei, ceea ce demonstrează că rolul statisticii nu se referă numai la cunoaşterea numerică a anumitor stări de fapte, dar că ea trebuie să fie folosită din plin şi în activitatea de transformare revoluţionară a societăţii. Aceasta presupune recunoaşterea caracterului partinic şi de clasă al statisticii social-economice, care se manifestă în interpretarea de pe poziţiile dialecticii materialiste a fenomenelor studiate.

V. I. Lenin arăta, în nenumărate rînduri, că fără statistică nu pot fi studiate realităţile social-economice. „O serie întreagă de probleme dintre cele mai esenţiale — scrie el — privind orînduirea economică a statelor moderne şi dezvoltarea ei, care înainte erau rezolvate pe baza unor consideraţii de ordin general şi a unor date aproximative, nu pot fi tratate în prezent într-un mod cît de cît de serios fără a lua în considerare datele mul-

54095

Scanned with OKEN Scanner

tiple referitoare la întregul teritoriu al țării respective, adunate conform unui anumit program și totalizate de specialiști statisticieni" [1].

Poziția fondatorilor socialismului științific este cît se poate de clară în această privință: cunoașterea vieții social-economice a unei orînduiri sociale nu este posibilă fără utilizarea statisticii. Ei au arătat că, cu cît procesul de producție se desfășoară pe scară socială mai largă, cu atît mai necesare devin evidența și statistica. În socialism, importanța statisticii a crescut în mod considerabil, asigurîndu-i-se condiții optime de dezvoltare. Nu se pot înfăptui conducerea științifică de stat și conducerea planificată a economiei socialiste fără cunoașterea desfășurării principalelor procese ale vieții economice, politice și culturale ale societății. Pentru aceasta este necesar să se cunoască adîncit cum este repartizată munca socială între ramurile economiei naționale, care este productivitatea muncii, ce fel de produse și în ce cantități se produc, în ce mod au loc schimbul și repartiția acestor produse etc.

Desigur, căile pentru cunoașterea acestor procese sînt multiple și variate. Una din căile principale, în condițiile economiei planificate, este aceea a folosirii datelor sistemului informațional statistic.

Statistica socialistă, bazîndu-se pe concepția materialist-dialectică, are un caracter științific fiind folosită pentru cunoașterea realității obiective, a mersului dezvoltării istorice a societății, a legilor economice obiective și a folosirii acțiunii lor pentru asigurarea creșterii continue a producției sociale, a nivelului de trai material și cultural al membrilor societății.

Pentru a înțelege rolul statisticii în procesul de cunoaștere a fenomenelor și proceselor social-economice este necesar să se precizeze, în primul rînd, sfera sa de cuprindere.

1.2. SFERA DE CUPRINDERE A STATISTICII SOCIAL-ECONOMICE

Plecînd de la clasificarea științelor potrivit principiilor filozofiei revoluționare a clasei muncitoare statistica face parte dintre acele ramuri ale științei care sînt scindate în două discipline aproape independente: una care se ocupă cu studiul fenomenelor din cadrul societății și alta care se ocupă cu studiul fenomenelor din natură și tehnologie. În această situație se găsesc pe lîngă statistică, geografia, care se împarte în geografia fizică și geografie economică, geologia în geologie fizică și geologie economică.

Clasificarea științelor amintite are la bază criterii obiective, bazate pe deosebirile calitative esențiale dintre diferitele forme de mișcare a materiei. Fenomenele sociale și economice, aparținînd unei forme superioare de mișcare a materiei, au un grad de complexitate mult mai mare și, deci, cunoașterea lor este cu atît mai complexă. De aceea, problemele de cunoaștere care se pun în fața statisticii social-economice sînt foarte variate și nuanțate și metodele de cercetare trebuie să fie și ele specifice. Astfel a apărut și s-a conturat ca o disciplină independentă statistica social-economică care aparține științelor sociale.

[1] V. I. Lenin, *Opere*, vol. 6, E.S.P.L.P., 1957, p. 425.

Scanned with OKEN Scanner

În general, obiectul de studiu al statisticii îl constituie fenomenele şi procesele în cadrul cărora este prezentă acţiunea legităţilor statistice, legităţi care se manifestă sub formă de tendinţă într-un număr mare de cazuri individuale diferite ca forme de manifestare, dar aparţinînd aceleiaşi esenţe. Astfel de fenomene se întîlnesc atît în natură, cît şi în societate şi exprimă rezultatul combinării mai multor factori determinanţi.

Examinarea fenomenelor şi proceselor sociale şi economice care formează obiectul de studiu al statisticii social-economice duce la concluzia că ea studiază o sferă foarte largă şi variată de fenomene ale vieţii sociale ce se referă la forţele şi relaţiile de producţie ale societăţii, cît şi la fenomene din viaţa politică şi culturală a societăţii.

Enumerarea cu titlu exemplificativ a unora dintre aceste fenomene şi procese demonstrează cu prisosinţă cîmpul foarte larg de cuprindere al obiectului statisticii sociale.

Statistica studiază mărimea şi structura avuţiei naţionale a unei ţări, în cadrul căreia se caracterizează mărimea şi structura fondurilor fixe din economia naţională, mărimea şi structura mijloacelor circulante pe ramuri şi forme de proprietate, acumulările de bunuri materiale ale populaţiei, resursele naturale ce pot fi atrase în procesul social al muncii, precum şi modificările intervenite de la o perioadă la alta.

Statistica studiază mărimea, structura, dinamica şi factorii de creştere a produsului social şi a venitului naţional, ca principale elemente ale realizării procesului reproducţiei socialiste lărgite.

Statistica studiază raporturile cantitative corespunzătoare unor relaţii calitative intervenite în variaţia productivităţii muncii, retribuirii după muncă, a costurilor şi a rentabilităţii.

Statistica studiază fenomenele demografice şi alte aspecte referitoare la populaţie, caracterizînd mişcarea populaţiei (natalitate, mortalitate etc.), nivelul ei de trai material şi cultural (retribuţia şi veniturile reale, consumul total şi consumul mediu pe un locuitor etc.); caracterizează statistic activitatea instituţiilor şi organizaţiilor social-culturale şi de servire a populaţiei.

Statistica studiază şi fenomene care fac parte din natură şi tehnologie, dar care influenţează asupra conţinutului şi dimensiunii unor procese sociale şi economice.

Statistica studiază fenomene din cadrul relaţiilor social-politice şi juridice ale societăţii.

Desigur, există fenomene sociale care nu fac parte din obiectul de studiu al statisticii, între care se pot aminti însuşirile artistice ale picturii, ale muzicii, ale literaturii etc.

Determinarea fenomenelor şi proceselor studiate de statistică, care au dus la precizarea că ea studiază societatea în întregul ei, nu a epuizat problema definirii şi delimitării obiectului său de studiu. Este necesar să se examineze care este raportul dintre statistica socială şi celelalte ştiinţe sociale, cum se delimitează obiectul său de studiu de al celorlalte ştiinţe sociale. Aici se pune, în primul rînd, problema raportului dintre statistica social-economică şi materialismul istoric şi economia politică marxistă.

Statistica, ca ştiinţă socială, se fundamentează pe ştiinţa celor mai generale legi ale dezvoltării societăţii omeneşti — materialismul dialectic

Scanned with OKEN Scanner

și istoric. La fel, statistica, al cărei obiect de studiu îl formează, în mare parte, fenomene și procese ale relațiilor de producție, se fundamentează ca știință pe economia politică marxistă.

Se poate spune, deci, că bazele teoretice ale statisticii sociale le constituie materialismul istoric și economia politică marxistă, deoarece ele sînt științele cele mai generale a căror obiect de studiu îl constituie cunoașterea legilor și elaborarea categoriilor economice și sociale cu care operează în mod unitar orice știință ce studiază fenomene din cadrul societății. În același timp, raportul dintre statistică și baza sa teoretică trebuie să fie înțeles ca un raport de interdependență și anume : statistica social-economică folosește în cercetarea sa concretă tezele și principiile elaborate de materialismul istoric și economia politică marxistă, pe care le ilustrează cantitativ, numeric în condiții concrete de timp și spațiu, furnizînd la rîndul ei material concret, documentar atît pentru verificarea principiilor teoretice, cît și pentru efectuarea unor noi generalizări.

Un raport de interdependență se realizează și între statistică și alte științe sociale, tehnice și ale naturii. Dată fiind baza sa teoretică, statistica trebuie să studieze fenomenele în strînsă interdependență cu dezvoltarea istorică, ceea ce implică folosirea tuturor cunoștințelor științifice din domeniul respectiv. În același timp, celelalte științe particulare nu se pot dispensa în studiile lor de bogatul material faptic pe care statistica îl poate pune la dispoziție și care nu poate fi prelucrat decît prin aplicarea metodelor statistice.

Astfel, de exemplu, studiind procesul reproducției socialiste lărgite, statistica trebuie să pornească de la tezele materialismului istoric, ale economiei politice marxiste, potrivit cărora modul de producție este factorul determinant, cauza dezvoltării societății iar mediul geografic și populația reprezintă condițiile materiale în cadrul cărora se desfășoară acest proces complex. De asemenea, va trebui să folosească cu competență categoriile economice — avuția națională, produsul social, venitul național, fondul de consum, fondul național de dezvoltare etc., precum și cunoștințele din domeniul geografiei economice, geologiei — pentru a putea analiza distribuția teritorială a forțelor de producție sau diferitele aspecte tehnologice, pentru a analiza eficiența introducerii tehnicii noi.

Deci, în cercetarea concretă statistica social-economică vine în contact cu diferite științe, fără a se suprapune obiectului acestora, ci numai prin folosirea reciprocă a rezultatelor obținute.

Această situație nu este proprie numai statisticii. Bogăția inepuizabilă a realității și complexitatea ei, cer eforturi conjugate multiple și din laturi variate de cunoaștere. De aceea trebuie examinate particularitățile statisticii social-economice, care o deosebesc de celelalte științe sociale în ceea ce privește obiectul său de studiu.

1.3. PARTICULARITĂȚI ALE OBIECTULUI DE STUDIU AL STATISTICII SOCIAL-ECONOMICE

Obiectul de studiu al statisticii, care permite delimitarea ei de celelalte științe, poate fi astfel formulată : statistica studiază *fenomenele sociale și economice de masă.*

Scanned with OKEN Scanner

Fenomenele din societate, tehnologice şi cele din natură pot să apară ca rezultat al influenţei unei singure cauze sau ca rezultat al influenţei combinate a unui număr mare de cauze.

Fenomenele din prima categorie apar ca fenomene simple, univoc determinate, care pe măsură ce se produce cauza se produce şi efectul, dacă condiţiile rămîn neschimbate. În cadrul acestor fenomene acţionează legi dinamice şi pentru a le cunoaşte şi caracteriza este suficientă efectuarea unor experimente, cu ajutorul cărora, odată legea descoperită, verificarea ei se poate face în fiecare caz în parte. În mod curent, despre aceste fenomene se vorbeşte ca despre fenomene deterministe a căror producere poate fi riguros stabilită.

Spre deosebire de aceste fenomene certe, ce se produc ori de cîte ori se produce cauza, se întîlnesc fenomene de masă care sînt fenomene complexe şi care apar ca rezultat al acţiunii combinate a unui număr mare de cauze şi condiţii. În cadrul acestor raporturi de cauzalitate, nu toate cauzele au acelaşi grad de esenţialitate şi acelaşi grad de influenţă. Mai mult, în procesul obiectiv de producere a acestor fenomene complexe acţiunea unora dintre factori se poate compensa, ca urmare a sensului diferit în care ei acţionează. Fenomenele social-economice sînt astfel de fenomene complexe, iar la analiza lor trebuie avut în vedere şi faptul că în cadrul mişcării sociale, în mod inevitabil influenţa factorilor obiectivi se asociază cu aceea a factorilor subiectivi.

Datorită faptului că, în mod obiectiv, fenomenele complexe depind de un număr mare de cauze şi condiţii, forma lor individuală de manifestare poate să difere foarte mult de la o unitate la alta, după modul concret de asociere a factorilor esenţiali cu cei întîmplători, aparent dînd impresia că ele s-ar produce la întîmplare, fără o cauzalitate certă. Raportate însă la întregul ansamblu, ele apar ca fiind asemănătoare între ele, generate de unele cauze esenţiale comune, diferenţiindu-se doar sub raport cantitativ. Evident, în acest caz, metoda de descoperire a legităţii de apariţie nu mai poate fi aceea bazată pe experimente (de laborator) ci pe metode mai analitice, bazate pe abstractizări şi generalizări succesive prin care se înlătură treptat ceea ce este neesenţial şi întîmplător în dezvoltarea acestor fenomene.

Spre deosebire de fenomenele ce se produc pe baza legilor dinamice şi care apar ca fenomene *identice*, fenomenele de masă sînt fenomene *asemănătoare*, a căror lege de apariţie nu poate fi cunoscută şi verificată decît la nivelul întregului ansamblu.

În literatura de specialitate, fenomenele de masă se mai numesc şi fenomene atipice, deoarece forma lor de manifestare individuală este diferită; sau fenomene de tip colectiv, deoarece legitatea de producere a lor nu poate fi cunoscută şi verificată în fiecare caz în parte, ci numai la nivelul întregului; sau fenomene de tip stohastic; sau indeterministe ca urmare a faptului că aici întîmplarea poate să modifice într-un sens sau altul tendinţa generală specifică întregului grup de fenomene din care fac parte. Rezultă că legităţile ce se manifestă în fenomenele şi procesele sociale de masă, nu pot fi scoase în relief, nu pot fi cunoscute şi caracterizate de statistică în schimbarea lor cantitativă, decît dacă se iau în studiu toate cazurile individuale sau o mare parte a lor, deoarece manifestările

Scanned with OKEN Scanner

cantitative individuale sînt extrem de variate. Astfel, pentru a putea cunoaşte nivelul de productivitate la care s-a ajuns într-o ramură de activitate, este necesar să se ia în considerare toate întreprinderile, cu toţi muncitorii care sînt ocupaţi în ramura respectivă şi nu cîţiva muncitori izolaţi.

Examinînd această particularitate a obiectului statisticii s-a abordat de fapt problema legii statistice, lege care se manifestă sub formă de tendinţă şi este valabilă numai pentru un ansamblu (o masă) de fenomene, nu pentru fiecare fenomen în parte. Este evident faptul că în cazul fenomenelor de masă — fenomene variabile între ele — descoperirea legilor care le determină, stabilirea raportului dintre fenomen şi esenţă, dintre particular şi general, se pot face numai prin aplicarea metodelor statistice. Specific acestor fenomene complexe sînt legile statistice, care, manifestîndu-se sub formă de tendinţă, pot fi cunoscute numai prin studierea unui mare număr de cazuri, în care abaterile individuale într-un sens sau altul sînt considerate ca fiind întîmplătoare şi se anihilează între ele, rămînînd reliefat numai ceea ce este comun şi tipic pentru întreaga masă.

Se ştie, de exemplu, că suma valorii mărfurilor coincide cu suma preţurilor lor, dar această lege nu poate fi descoperită şi nici verificată în cazul fiecărei mărfi în parte. V.I. Lenin arăta: „. . . este cu totul firesc ca într-o societate de producători de marfă, răzleţi, care sînt legaţi între ei numai prin piaţă, legile necesare să nu se poată manifesta decît sub formă de legi medii, de legi sociale, de legi ale unor fenomene de masă, abaterile individuale într-o parte sau în altă parte compensîndu-se reciproc” [1].

Legea statistică poate fi, deci, descoperită numai prin cercetarea unui număr foarte mare de cazuri individuale întîmplătoare, legate între ele prin acţiunea comună a aceloraşi cauze obiective. Caracteristicile legii — obiectivitate, necesitate, generalitate, repetabilitate — se întîlnesc şi în cazul legilor statistice. Rezultă că cercetarea statistică nu se poate efectua decît dacă se identifică întregul grup de fenomene asemănătoare ce au fost generate de aceeaşi lege, în cadrul căruia acţiunea factorilor întîmplători este compensată reciproc. Pentru aceasta este necesar ca în procesul de cercetare concretă să se pornească de la depistarea tuturor cazurilor individuale şi abstractizînd succesiv să se elimine ceea ce este întîmplător în producerea fenomenelor, desprinzînd tendinţa lor obiectivă de dezvoltare pe care să o exprime prin mijloace statistice. În acest sens trebuie înţeleasă cercetarea statistică, în cadrul căreia fenomenele studiate sînt interpretate ca fenomene probabile, cu grad mare de variaţie de la o unitate la alta, dar care aparţin aceleiaşi esenţe social-economice. Aceasta impune în mod necesar ca în cercetarea statistică să se ţină seama de principiile teoriei probabilităţii şi de cerinţele legii numerelor mari.

Deci, legea numerelor mari — considerată indispensabilă pentru interpretarea legii statistice — presupune că într-un număr suficient de mare de cazuri individuale, a căror variaţie este întîmplătoare, influenţa factorilor într-un sens sau altul se compensează reciproc în aşa fel încît se obţine o anumită valoare tipică, caracteristică întregului ansamblu.

---

[1] V. I. Lenin, *Opere*, vol. 21, E.S.P.L.P., 1958, p. 53.

Scanned with OKEN Scanner

Dacă, însă, ansamblul de fenomene de masă este format din mai multe tipuri calitative, atunci înseamnă că şi factorii esenţiali sînt variabili între ei, de la un tip calitativ la altul. Pentru astfel de fenomene compensarea abaterilor se face atît pe întregul ansamblu, cît şi pentru fiecare tip calitativ în parte.

De reţinut este faptul că, în general, toate fenomenele, deci cu atît mai mult fenomenele sociale şi economice, nu se supun numai legilor statistice sau legilor dinamice, ci în fiecare dintre ele se împleteşte acţiunea mai multor legi diferite. Acest fapt este deosebit de important pentru statistica social-economică, unde este mai necesar poate decît în alte statistici, să se înţeleagă limitele legii statistice şi ale legii numerelor mari în explicarea fenomenelor sociale şi economice.

O altă particularitate a statisticii ca ştiinţă socială constă în faptul că ea studiază *aspectul cantitativ concret al fenomenelor social-economice de masă, în condiţii specifice* de timp şi loc.

Deci, specific statisticii este faptul că ea cuprinde în obiectul său de studiu fenomenele social-economice de masă sub aspectul manifestării lor cantitative. Fără îndoială, aici trebuie reamintită concepţia filozofiei marxiste potrivit căreia latura cantitativă a fenomenelor nu se prezintă separat de calitate, cantitatea neexistînd decît ca o cantitate a unei calităţi date.

Aceasta înseamnă că, cercetînd statistic fenomenele, accentul trebuie să cadă pe analiza laturii cantitative, fără a putea ignora însă aspectul lor calitativ.

Deci, numai bazîndu-se pe cunoaşterea legilor vieţii sociale statistica poate studia manifestarea lor concretă în timp şi spaţiu. Marx arăta limpede că „Statistica ajunge în stare să treacă la analiza reală a ratei salariului în diferite epoci şi ţări abia după ce desluşeşte raporturile care formează profitul" [1].

În acelaşi timp, statistica procedează la cunoaşterea fenomenelor sociale concrete prin abstractizare şi generalizare, trecînd de la date individuale numeroase, variate ca formă de manifestare, la elaborarea unui sistem de indicatori, care este conceput ca un sistem de noţiuni, de categorii (medie, ritm de creştere etc.) corespunzător trăsăturilor esenţiale ale fenomenelor cercetate, pentru fiecare etapă în parte, în condiţii specifice de spaţiu. Indicatorii statisticii apar astfel ca o *generalizare* a laturii cantitative a fenomenelor sociale şi economice în strînsă legătură cu latura lor calitativă. Reflectarea vieţii sociale şi economice cu ajutorul indicatorilor statisticii în fiecare etapă dată constituie conţinutul principal al muncii practice statistice.

1.4. METODA STATISTICII

După cum se ştie, metoda de cercetare a unei ştiinţe depinde, pe de o parte, de baza sa teoretică şi, pe de altă parte, de obiectul său de studiu. Întrucît baza teoretică a statisticii este materialismul istoric şi economia politică marxistă, iar obiectul ei îl constituie manifestările cantitative ale

[1] K. Marx, *Capitalul*, vol. III, partea I, E.S.P.L.P., 1956, p. 240–241.

Scanned with OKEN Scanner

fenomenelor de masă din cadrul vieţii sociale, înseamnă că metoda ei trebuie să corespundă principiilor bazei sale teoretice şi conţinutului fenomenelor pe care le studiază.

Potrivit bazei sale teoretice, statistica foloseşte în procesul de cunoaştere metoda materialist-dialectică. Acest lucru nu exclude, ci dimpotrivă, impune ca statistica, la fel ca celelalte ştiinţe, să posede o metodă proprie de cercetare în funcţie de particularităţile obiectului său de studiu.

Potrivit obiectului său de studiu, statistica şi-a elaborat procedee de cercetare, cum sînt cele ale observării de masă, ale centralizării şi grupării, procedee pentru determinarea indicatorilor generalizatori, procedee şi modele de analiză şi interpretare statistică, care constituie în totalitatea lor metoda statisticii.

Urmînd fazele cercetării — corespunzătoare dialecticii procesului de cunoaştere în general — se poate spune că metoda statisticii constă în totalitatea procedeelor, tehnicilor şi principiilor utilizate pentru efectuarea observării fenomenelor sociale de masă, pentru prelucrarea datelor obţinute prin observare şi pentru analiza şi interpretarea rezultatelor statistice obţinute.

Procedeele de observare statistică creează posibilitatea cunoaşterii formelor individuale multiple şi variate ale fenomenelor studiate, în manifestarea lor cantitativă. Observarea statistică se realizează prin procedee specifice sub formă de dări de seamă, recensăminte, anchete, observări selective, monografii. Rezultatele observării se prezintă ca o masă de cunoştinţe variate despre felul în care fenomenul sau procesul se manifestă în fiecare caz în parte. În acest stadiu încă nu se cunoaşte şi nu se poate caracteriza numeric esenţa fenomenului.

Materialul obţinut prin observare este supus, cu ajutorul unor procedee specifice statisticii, unor prelucrări succesive cînd ceea ce este întîmplător în manifestările individuale se elimină şi se păstrează ceea ce este comun şi esenţial, în expresii numerice generalizatoare. În cadrul acestei etape îşi găsesc o largă aplicabilitate : metoda grupării, metoda mediilor, analiza dispersională, analiza corelaţiei, metoda indicilor şi altele.

Aplicarea acestor metode şi procedee are ca rezultat obţinerea sistemului de indicatori format din mărimi : absolute, relative, medii, indicatori de variaţie şi corelaţie, indici, ecuaţii de estimare a tendinţelor etc.

În sfîrşit, procedeele de interpretare şi analiză a rezultatelor cercetării permit închegarea procesului de cunoaştere statistică, prin examinarea expresiilor numerice statistice generalizatoare, obţinute prin prelucrare în spiritul legilor obiective de dezvoltare a societăţii.

Aplicarea metodei statisticii sociale trebuie să se facă diferenţiat în raport cu natura specifică a fenomenelor cercetate, cu scopul cercetării şi cu tehnicile de calcul existente.

Procedeele de observare, prelucrare, analiză şi interpretare a datelor statistice formează conţinutul cursului de „Statistică teoretică" sau „Metoda statisticii social-economice" sau „Statistică generală", cum i se mai spune, şi ele se aplică în mod unitar pentru toate disciplinele statistice, care studiază fenomenele şi procesele din diferitele ramuri ale economiei naţionale (industria, agricultura, comerţ, nivel de trai etc.) sau din perspectiva întregii economii (statistica economiei naţionale).

Scanned with OKEN Scanner

Aceste discipline statistice s-au diferenţiat ca urmare a adîncirii procesului de diviziune socială a muncii, care face ca problemele pe care le cercetează statistica să fie din ce în ce mai complexe şi mai variate. Indiferent de ramura de activitate sau procesul studiat, ramurile statisticii fac parte prin obiect şi metodă din statistica social-economică, diferenţa specifică dintre ele fiind determinată de conţinutul social-economic particular al fiecărei ramuri. Caracterul unitar al acestor discipline relativ independente ale statisticii social-economice derivă şi din faptul că toate folosesc aceleaşi noţiuni şi metode.

1.5. PRINCIPALELE NOŢIUNI FOLOSITE ÎN STATISTICA SOCIAL-ECONOMICĂ

Printre aceste noţiuni se pot aminti : colectivitatea statistică, unitatea statistică, caracteristica statistică şi indicatorul statistic.

Colectivitatea [1] statistică reprezintă totalitatea cazurilor individuale de acelaşi fel, formate pe baza influenţei aceloraşi cauze esenţiale şi care constituie obiectul supus cercetării. Statistica social-economică, studiind fenomene din cadrul vieţii sociale, se ocupă cu colectivităţi statistice cu caracter obiectiv şi finit. Prin aceasta ea se deosebeşte de statistica matematică, care, de regulă, se ocupă de colectivităţi statistice abstracte şi infinite.

Caracterul obiectiv al colectivităţii statistice este determinat de faptul că fenomenele studiate se produc independent de voinţa cercetătorului şi el are doar sarcina de a le identifica şi măsura statistic, iar aplicînd metodele cele mai adecvate, poate trece la formularea unor concluzii statistice care să reflecte imaginea reală a manifestării acestor fenomene.

În acelaşi timp, fenomenele sociale avînd un caracter concret, colectivitatea statistică apare ca un ansamblu finit de unităţi, măsurabil printr-un procedeu de observare statistică adecvat.

Ţinînd seama de caracterul obiectiv şi finit al colectivităţii statistice din domeniul social şi economic, apare necesitatea ca de fiecare dată cînd se organizează o cercetare statistică să se delimiteze colectivitatea statistică în timp şi spaţiu. Pentru aceasta este necesar să se efectueze o analiză teoretică prealabilă în scopul stabilirii unor criterii obiective potrivit cărora să se facă identificarea tuturor cazurilor asemănătoare care formează colectivitatea ce trebuie studiată.

Un exemplu de colectivitate statistică este populaţia, care este numeroasă, prezintă un grad mare de variaţie a formelor individuale de manifestare; ea constituie atît obiectul de cercetare al statisticii demografice cît şi al cercetărilor concrete sociologice.

Unitatea statistică reprezintă forma individuală de manifestare obiectivă a fenomenelor supuse cercetării. Unităţile statistice sînt elementele constitutive ale colectivităţii statistice. Unităţile statistice pot fi unităţi simple sau complexe.

---

[1] În statistica matematică „populaţie".

Scanned with OKEN Scanner

Unităţile statistice simple ţin de modul obiectiv de apariţie a fenomenelor sociale şi reprezintă forma concretă, nemijlocită, sub care se întîlnesc fenomenele în societate (persoana în cadrul populaţiei, muncitorul în cadrul industriei etc.).

Unităţile statistice complexe depind de modul de organizare social-economică a fenomenelor şi proceselor, cum sînt: familia în cercetarea demografică şi sociologică, echipa, secţia, întreprinderea în cadrul unei ramuri de producţie etc. Unităţile statistice complexe sînt formate din mai multe unităţi simple şi culegerea datelor trebuie să se facă pentru toate aceste unităţi care formează împreună o unitate statistică complexă.

Caracteristica statistică reprezintă însuşirile fenomenelor studiate. Unităţile unei colectivităţi statistice sînt purtătoare ale unor caracteristici variabile ca formă de manifestare sau ca nivel de dezvoltare. De aceea, caracteristicile statistice se mai numesc şi variabile statistice sau variabile aleatoare, iar formele concrete de manifestare ale acestor caracteristici se numesc variante. Într-o cercetare statistică variantele se pot întîlni o singură dată sau de mai multe ori. Numărul de unităţi la care se înregistrează aceeaşi variantă se numeşte frecvenţă sau pondere.

Caracteristicile statisticii sînt de mai multe feluri:

— după conţinutul caracteristicii
  - caracteristici de timp (exemplu: anul înfiinţării unei întreprinderi);
  - de spaţiu (de exemplu: judeţul de care aparţine);
  - atributive — orice caracteristică care se poate exprima numeric sau prin cuvinte;

— după natura variaţiei
  - caracteristici cu variaţie continuă (cele care pot lua orice valori într-un interval dat; exemplu: vîrsta);
  - caracteristici cu variaţie discontinuă sau discreată (cele care nu pot lua decît valori dispuse la anumite intervale, de exemplu: numărul de muncitori pentru o întreprindere, mărimea fondurilor fixe etc.);

— după modul de obţinere şi caracterizare a fenomenului
  - caracteristici primare, întîlnite la toate unităţile simple ale colectivităţii;
  - caracteristici derivate întîlnite la nivelul unităţilor complexe, determinate prin compararea a doi indicatori primari. În cercetarea statistică caracteristicile derivate se folosesc ca şi cele primare, iar frecvenţele lor sînt numitorii rapoartelor;

— după forma de manifestare la nivelul unităţilor simple
  - caracteristici alternative, cele care au doar două variante; de exemplu, producţie bună sau rebut;
  - caracteristici nealternative sînt toate celelalte caracteristici, ca de exemplu: virsta, profesia, calificarea etc.

Scanned with OKEN Scanner

Indicatorii statistici reprezintă expresia numerică a unei determinări calitative obiective, obținută în urma efectuării unei cercetări statistice, raportată la anumite condiții specifice de timp și spațiu. Deci, indicatorul statistic cuprinde atît noțiunea care determină conținutul indicatorului, cît și expresia sa numerică raportată la anumite condiții de timp și spațiu. Indicatorii statistici pot să se refere la caracteristicile statistice studiate sau la ilustrarea numerică a diferitelor categorii economice. Sarcina statisticii este de a elabora un sistem de indicatori statistici care să permită caracterizarea statistică și dinamică a principalelor aspecte ale dezvoltării vieții social-economice.

Fenomenele studiate de statistică, fiind variabile între ele, nu pot fi caracterizate printr-un singur indicator. De aceea, în orice cercetare statistică se folosește un sistem de indicatori care exprimă numeric multitudinea relațiilor cantitative dintre fenomenele studiate.

Caracterizarea statistică a fenomenelor și proceselor de tip stohastic nu se poate face decît după principiile cercetării sistemice, folosind relațiile de interdependență dintre părțile componente și întregul ansamblu de fenomene. Aceasta conduce de fapt la elaborarea unor modele de analiză statistică care să fie aplicată la nivel micro- și macroeconomic, folosind datele din sistemul informațional economic.

Sistemul de indicatori statistici asigură însuși conținutul muncii statistice și este elementul care unește cele trei faze ale cercetării statistice (observare, prelucrare, analiză).

O problemă de bază a elaborării sistemului de indicatori statistici este aceea a definirii conținutului indicatorilor și a stabilirii nomenclatoarelor de clasificare a acestora în funcție de necesitățile de informație ale organelor de conducere și decizie.

În socialism, sistemul de indicatori statistici se elaborează în funcție de sistemul de indicatori ai planului. În consecință, fiecărui indicator de plan îi corespunde un indicator statistic, care trebuie să fie determinat după aceleași principii organizatorice și metodologice. Statistica, urmărind nu numai gradul ci și modul de realizare a sarcinilor planificate în scopul descoperirii factorilor care au influențat și a rezervelor neutilizate, se bazează pe un sistem mai larg de indicatori decît cel al planului și cuprinde în sfera sa de cunoaștere întreaga viață economică și social-culturală.

1.6. ORGANIZAREA CERCETĂRII STATISTICE

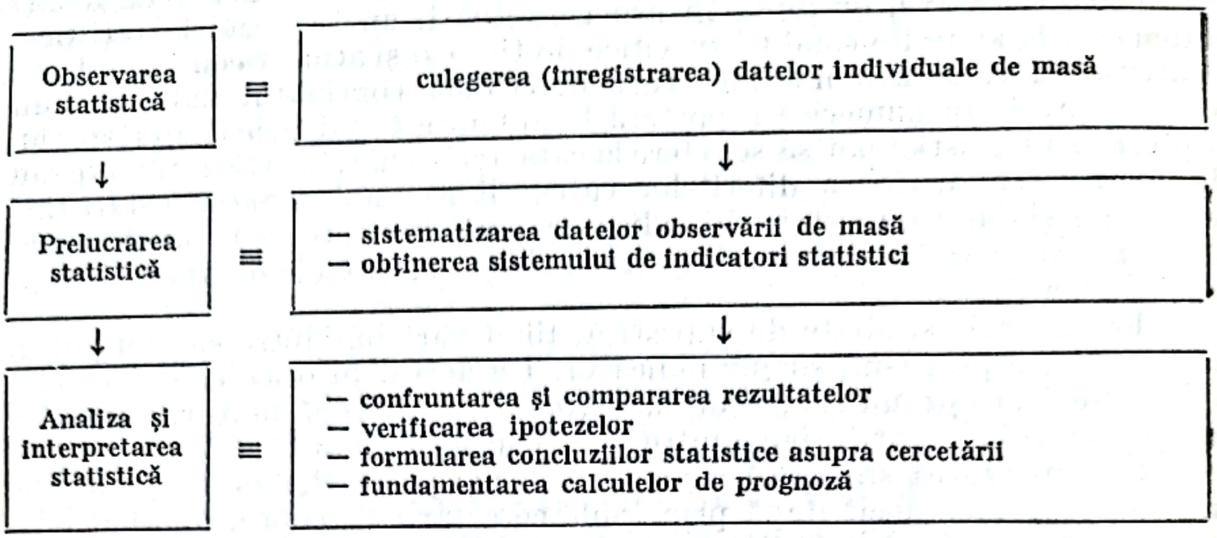
După cum s-a arătat, cercetarea statistică se desfășoară în trei etape succesive : observarea, prelucrarea și analiza statistică.

Într-o formă generală, procesul de cunoaștere statistică se prezintă ca în schema I.

Deși cercetarea statistică se desfășoară în aceste trei etape succesive, ea reprezintă un proces unitar prin conținutul, scopul și prin metodele utilizate. În cadrul cercetării statistice rolul cel mai important îi revine analizei statistice care trebuie să asigure nu numai o interpretare științifică a rezultatelor cercetării, dar mai ales obținerea unui material documentar real, prelucrat după metodele cele mai adecvate naturii specifice fenomenelor cercetate.

Scanned with OKEN Scanner

Etapele cercetării statistice



Schema I.

Procesul de cercetare statistică se poate realiza în condiții optime folosind echipamentele moderne de calcul, care prezintă un dublu avantaj : se pot culege date pentru un număr mai mare de caracteristici și se reduce foarte mult timpul de prelucrare a informațiilor culese.

În general, cercetările social-economice sînt operațiuni de mare amploare și ele trebuie pregătite cu multă atenție și competență. De aceea, cercetarea statistică trebuie să se bazeze pe un plan dinainte stabilit, în care să se precizeze atît măsurile organizatorice, cît și problemele cu caracter metodologic ce trebuie rezolvate în efectuarea concretă a cercetării.

Cercetarea statistică trebuie privită ca un proces complex, a cărui formă ciclică poate fi prezentată în următoarea schemă (vezi schema II).

Toate aceste elemente se regăsesc în planul cercetării statistice care cuprinde problemele ce trebuie rezolvate în fiecare fază a observării, a prelucrării, precum și a analizei și interpretării rezultatelor. Pentru aceasta este necesar ca înainte de a trece la elaborarea planului să se facă o documentație amplă asupra modului în care astfel de cercetări au fost organizate în trecut la noi în țară sau pe plan mondial, asupra posibilităților concrete de folosire a surselor de informații necesare.

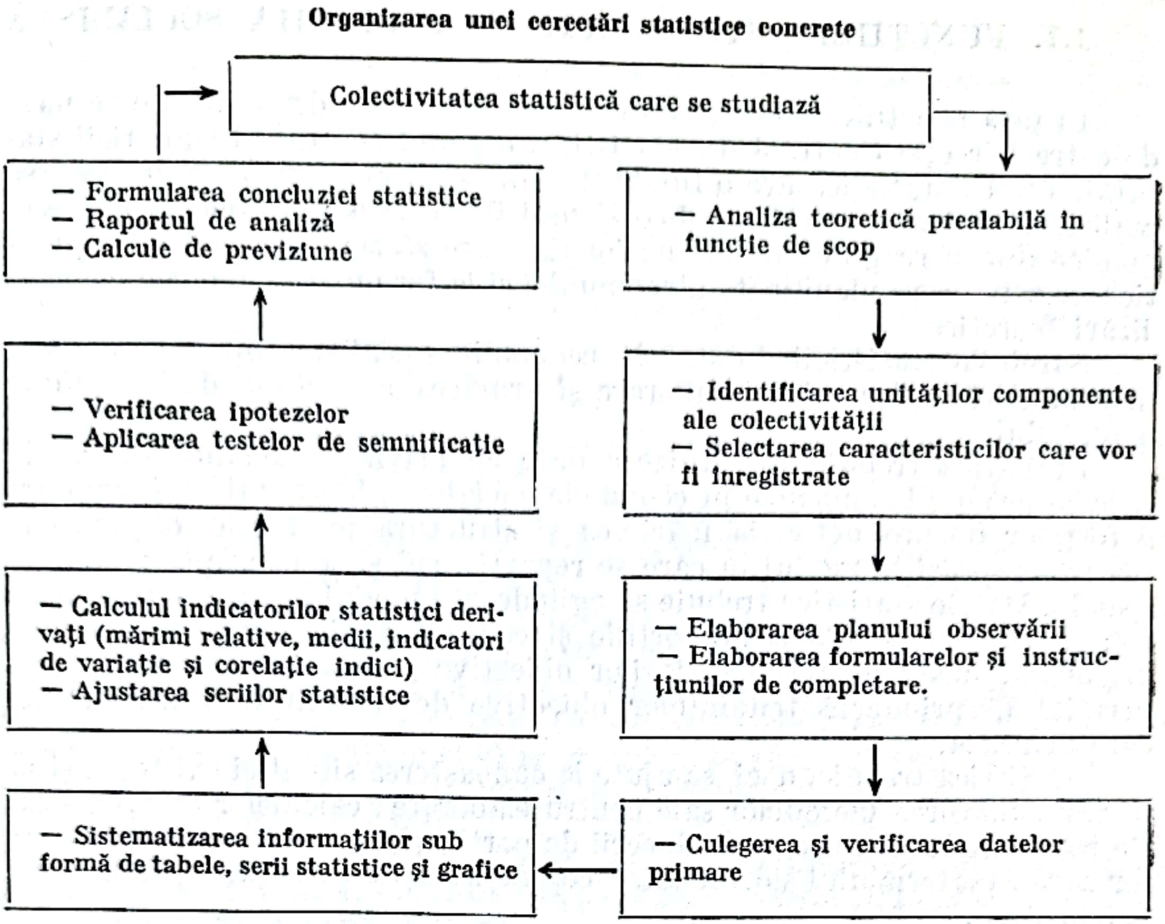
De cele mai multe ori ordinea de elaborare a planului cercetării este inversă efectuării procesului de cercetare : întîi se stabilește ce anume se va analiza și în ce scop, în funcție de care se elaborează sistemul de indicatori și metodele de prelucrare ce vor fi folosite și, în sfîrșit, pe baza acestora se aleg caracteristicile ce trebuie înregistrate și forma de observare. Indiferent de ordinea în care se elaborează planul, elaborarea lui este subordonată obiectului și scopului cercetării și trebuie făcută pentru fiecare cercetare în parte.

În societatea socialistă cercetarea statistică se poate realiza pe două căi :

— calea prin care în toate etapele cercetării se folosesc metode și procedee statistice și în acest caz apar cercetări statistice special organizate și care se realizează prin recensăminte, observări selective, anchete sau monografii statistice. Datele culese pe această cale statistică sînt pre-

Scanned with OKEN Scanner

Organizarea unei cercetări statistice concrete



Schema II.

lucrate şi analizate pentru a putea caracteriza multilateral sub raport statistic fenomenele cercetate;

— cea de-a doua cale, în care statistica foloseşte combinat date şi din alte forme de evidenţă (contabilă, tehnico-operativă). În acest caz datele prelucrate şi analizate statistic sînt culese şi înregistrate folosind întregul sistem informaţional al economiei naţionale. Aceasta este calea prin care se întocmesc dările de seamă care servesc ca mijloc de înregistrare curentă pentru analiza modului de realizare a procesului reproducţiei socialiste lărgite la fiecare etapă dată.

Existenţa proprietăţii socialiste asupra mijloacelor de producţie în economia naţională face necesară şi posibilă organizarea unui sistem informaţional unitar, care să permită centralizarea datelor din toate unităţile economice cu privire la activitatea curentă de desfăşurare a îndeplinirii sarcinilor de plan. Unitatea lor este asigurată de folosirea unor principii unice cu privire la organizarea şi metodologia de calcul. Sistemul informaţional economic şi social se bazează pe datele evidenţei tehnico-operative, contabile şi statistice. Sursa comună de informare o constituie documentaţia primară.

Aceste elemente componente ale sistemului informaţional formează un tot unitar care asigură în mod permanent culegerea, prelucrarea, stocarea şi transmiterea informaţiei necesare conducerii ştiinţifice a economiei naţionale pe diferitele sale trepte de decizie.

Scanned with OKEN Scanner

1.7. FUNCȚIILE STATISTICII ÎN ECONOMIA SOCIALISTĂ

În țara noastră, munca practică statistică este îndrumată și efectuată de către Direcția Centrală de Statistică, organul central al statisticii statului. Munca statistică are o finalitate bine stabilită, orientată în vederea realizării unei cunoașteri veridice și multilaterale a proceselor social-economice din întreaga economie națională, care să servească nevoilor practice de conducere planificată a economiei și la formularea unor noi generalizări teoretice.

Sarcinile statisticii de stat în economia socialistă sînt subordonate, în primul rînd, muncii de elaborare și urmărire a modului de îndeplinire a planului.

Statistica trebuie să furnizeze date cu privire la modul concret în care se produc fenomenele în economia socialistă, la gradul de dezvoltare a forțelor de producție, la mărimea și structura producției obținute în fiecare etapă și la modul în care se repartizează și se consumă producția socială. Datele statistice trebuie să oglindească modul în care se realizează concret în fiecare etapă proporțiile și corelațiile din întreaga economie națională, ilustrînd acțiunea legilor obiective economice și sociale și să permită desprinderea tendințelor obiective de dezvoltare pentru etapa care urmează.

Statistica trebuie, deci, să ajute la cunoașterea situației existente și să permită folosirea metodelor sale pentru elaborarea calculelor de prognoză. Pe baza datelor furnizate conducerii de partid și de stat, ea constribuie la formarea materialului documentar ce stă la baza elaborării planului de perspectivă.

În cadrul unei economii planificate statistica are funcția de a urmări modul de realizare a sarcinilor de plan. Analiza modului de îndeplinire a planului presupune descoperirea raporturilor de cauzalitate în vederea determinării gradului de influență a diferiților factori, precum și evidențierea rezervelor neutilizate în perioada respectivă.

Problemele din ce în ce mai complexe, pe care practica socialismului le ridică, impun statisticii social-economice o dezvoltare și perfecționare continuă a metodelor și procedeelor de observare, prelucrare și analiză, o adaptare continuă a lor la condițiile moderne de calcul. Dotarea economiei naționale cu echipamente moderne de calcul, înființarea centrelor teritoriale de calcul pun în fața statisticii sarcini multiple legate de perfecționarea sistemelor de indicatori, a clasificării și elaborării modelelor de prelucrare statistică, a tipizării formularelor în vederea asigurării unei creșteri calitative a întregii informații statistice. Acestea au ca rezultat nu numai lărgirea ariei de investigație statistică, dar mai ales adîncirea muncii de analiză statistică. Aplicarea mai largă a instrumentelor moderne de calcul și a metodelor și procedeelor statisticii matematice au ca rezultat cunoașterea într-o mai mare măsură și mai adîncită a raporturilor cantitative dintre fenomenele cercetate, a mutațiilor produse în structura socială a economiei naționale, permit adîncirea analizelor comparative pe plan internațional.

CAPITOLUL II

OBSERVAREA STATISTICĂ

Realizarea cunoaşterii statistice a realităţii social-economice presupune ca în prima etapă să se culeagă informaţii despre aspectele sub care se prezintă fenomenele. Această etapă este cunoscută sub denumirea de observare statistică sau înregistrare statistică.

Specificul obiectului de cercetare al statisticii determină unele particularităţi ale observării.

Observarea statistică trebuie făcută la fiecare unitate a colectivităţii în parte. Cum numărul unităţilor este foarte mare şi ele sînt răspîndite pe o suprafaţă întinsă, înregistrarea simultană a lor nu se poate face decît folosind un număr mare de observatori, care să aplice în mod unitar aceleaşi criterii de culegere a datelor.

Ţinînd seama că faptele şi procesele supuse cercetării statistice se găsesc în interdependenţă cu alte fapte şi fenomene, cu care uneori se pot confunda uşor, este necesar să se stabilească criterii sigure pentru a identifica numai unităţile care aparţin colectivităţii studiate. Pentru aceasta este necesar să se identifice şi să se înregistreze acele însuşiri ale fenomenelor care sînt comune şi esenţiale pentru toate unităţile colectivităţii şi corespund scopului cercetării.

Datele obţinute prin cercetarea statistică trebuie să îndeplinească atît condiţia de volum, cît şi de calitate. Dacă datele sînt incomplete ca volum, nu există siguranţa că prin prelucrarea lor, oricît de corectă ar fi, se va ajunge la descoperirea legităţilor obiective după care s-au produs fenomenele. În acelaşi timp trebuie să fie luate toate măsurile ca datele observării să nu conţină erori, pentru ca valoarea indicatorilor generalizatori obţinuţi pe baza lor să nu fie modificată, iar concluzia statistică să reflecte denaturat fenomenele studiate. Numai astfel este posibil să se obţină date cu conţinut real pe întreaga filieră a procesului de cercetare.

Rezultă că observarea constituie etapa cercetării statistice în care se culeg, după criterii unitare, informaţii despre formele individuale de manifestare ale fenomenelor sociale şi economice de masă. Informaţiile culese (datele statistice) trebuie să se prezinte ca un material informativ complet şi unitar atît sub aspectul organizatoric, cît şi al conţinutului. Observarea statistică urmăreşte deci stabilirea unui fundament de fapte concrete pe baza cărora se pot face generalizările necesare explicării ştiinţifice a fenomenelor şi proceselor studiate, prin descoperirea legităţii după care ele se produc. În acest sens ea trebuie privită ca o operaţie dependentă de celelalte faze ale procesului de cunoaştere statistică.

2.1. ORGANIZAREA UNEI OBSERVĂRI STATISTICE

Ca etapă a cercetării, observarea statistică este condiţionată de scopul cercetării, de particularităţile obiectului studiat şi de tehnica de calcul folosită în procesul de prelucrare ulterioară a datelor observării.

Scanned with OKEN Scanner

În funcţie de aceşti factori se organizează observarea care presupune fie culegerea directă a datelor pe căi statistice, fie preluarea şi înregistrarea în formulare statistice a unor date deja existente în diferite documente.

În economia socialistă, existenţa sistemului informaţional economic permite lărgirea sferei de cuprindere a observării statistice cu caracter permanent. Pe lîngă aceasta, ori de cîte ori este nevoie să se obţină un surplus de informaţie statistică, se organizează cercetări speciale, care presupun folosirea unor forme specifice de observare statistică.

Pentru organizarea raţională a unei observări statistice trebuie rezolvate o seamă de probleme, unele cu caracter metodologic şi altele cu caracter organizatoric. Principalele probleme care trebuie rezolvate pentru realizarea unei observări fundamentate, care constituie de fapt tematica planului oricărei observări statistice, se referă la : scopul observării ; obiectul observării ; programul observării ; unitatea de observare ; timpul şi locul observării ; formularele, instrucţiunile ; problemele organizatorice necesare desfăşurării normale a operaţiei de culegere a datelor.

Scopul observării este subordonat scopului general pentru care s-a organizat cercetarea statistică.

În economia socialistă, scopul general al cercetării este legat de funcţiile statisticii ca ştiinţă socială, de necesităţile de conducere planificată a economiei naţionale, de nevoile de informaţii pentru luarea deciziilor pe toate treptele organizatorice. Scopul special al unei observări se determină în funcţie de necesităţile specifice de informaţii pentru domeniul în care se organizează cercetarea şi care se încadrează în scopul general al acesteia.

Stabilirea precisă a scopului cercetării şi, respectiv, al observării are mare importanţă practică pentru rezolvarea unor probleme, cum ar fi delimitarea obiectului de observare, definirea unităţilor de observare, stabilirea programului observării.

Definirea clară şi concisă a scopului observării permite selectarea caracteristicilor care sînt strict necesare, precum şi alegerea celor mai potrivite procedee de culegere a datelor.

Obiectul observării îl formează colectivitatea de fenomene despre care urmează să se culeagă date statistice.

De remarcat este faptul că nu întotdeauna obiectul cercetării coincide cu obiectul observării. Astfel, în cazul cînd se foloseşte o observare parţială, colectivitatea observată este mai mică decît colectivitatea totală. Indiferent de felul observării utilizate, este necesar ca atunci cînd caracterizarea finală se referă la întreaga colectivitate să se definească nu numai colectivitatea parţială — pentru care se culeg datele — ci şi cea totală — pentru care se generalizează rezultatele. Definirea colectivităţii statistice trebuie să se facă după o serie de caracteristici esenţiale, stabile şi care să asigure comparabilitatea datelor de la o observare la alta.

Programul observării este format din totalitatea caracteristicilor pentru care urmează să se culeagă date de la toate unităţile incluse în obiectul observării.

Întreaga muncă de cercetare statistică este condiţionată de identificarea corectă a acelor caracteristici care sînt strict necesare elaborării sistemului de indicatori.

Scanned with OKEN Scanner

Programul observării se concretizează în întrebări, care sînt înscrise în formulare statistice şi la care trebuie să se obţină răspunsuri cu privire la nivelul de dezvoltare a caracteristicilor sau la forma lor de manifestare.

Unitatea de observare reprezintă elementul constitutiv al colectivităţii pentru care s-a organizat o observare statistică. În funcţie de scopul cercetării se pot folosi unităţi de observare simple şi complexe. De obicei, în observare unităţile simple se înregistrează în cadrul unităţilor complexe din care fac parte.

Unităţile de observare statistică, în special cele complexe, trebuie să fie definite de fiecare dată cînd se organizează o culegere de date, folosind caracteristici stabile şi care sînt comune tuturor unităţilor.

Definirea unităţilor de observare trebuie să se facă pe baza unor elemente clare, care să permită înţelegerea unitară a noţiunii respective de către toţi cei care culeg date. De regulă, definirea unităţii trebuie să fie însoţită de exemple de manifestare concretă sub care se pot întîlni diferite cazuri, în special, acelea care se rezolvă pe baza unor criterii convenţionale.

La definirea unităţii de observare trebuie să se ţină seama şi de necesitatea de a asigura comparabilitatea datelor pe plan teritorial şi în dinamică.

Timpul şi locul observării. Pentru efectuarea unitară a observării statistice trebuie să se precizeze atît timpul la care se referă datele culese, cît şi timpul cînd se face înregistrarea.

Timpul la care se referă datele nu trebuie confundat cu timpul înregistrării. De regulă, se spune că timpul observării este acela la care se referă datele culese. Acesta poate să fie un singur moment (pentru înregistrările statice) sau o perioadă (pentru înregistrările dinamice).

De exemplu, la ultimul recensămînt al populaţiei din ţara noastră, momentul critic (momentul la care se refereau datele culese) a fost ora „0" din noaptea de 4 spre 5 ianuarie 1977, iar timpul de înregistrare a datelor a fost perioada 5—12 ianuarie 1977. Deci timpul observării, în acest exemplu, îl constituie momentul critic. Indiferent de situaţia existentă în timpul înregistrării, recenzorul a consemnat situaţia care corespundea momentului critic. Dacă, de pildă, o persoană decedase în ziua de 7 ianuarie, cînd recenzorul a efectuat observarea, ea a fost trecută ca fiind în viaţă; un copil care s-a născut după momentul critic nu a fost înregistrat. În acest fel se asigură unitatea tuturor datelor înregistrate pe întreg teritoriul ţării, ca şi cînd s-ar fi efectuat o înregistrare simultană a populaţiei.

Pentru înregistrările în dinamică se stabileşte timpul observării pentru o perioadă — o zi, o decadă, o lună, un trimestru etc. De exemplu, darea de seamă pe luna ianuarie se completează în primele zile ale lunii februarie, după ce a expirat timpul la care se refereau datele.

Există şi situaţii cînd timpul observării coincide cu timpul cînd se înregistrează datele, cum este cazul fotografierii zilei de lucru a muncitorului.

În prelucrarea ulterioară a datelor nu apare decît timpul observării, adică timpul la care se referă datele.

Un alt aspect al problemei este alegerea timpului observării. În practica statistică, ca timp al observării se alege perioada în care colectivitatea

Scanned with OKEN Scanner

prezintă gradul cel mai ridicat de stabilitate. De exemplu, dacă se organizează un sondaj pentru cunoaşterea elasticităţii cererii generale de mărfuri trebuie să se evite să se ia ca perioadă de înregistrare zilele apropiate de sărbătorile de iarnă care modifică substanţial volumul şi structura cererii de mărfuri. De asemenea pentru recensămintele populaţiei se preferă perioada de iarnă cînd se înregistrează cea mai mare stabilitate.

Locul unde se înregistrează datele coincide, de regulă, cu locul producerii fenomenelor. Cazurile în care locul înregistrării diferă de locul producerii fenomenelor se întîlnesc în special cînd observarea statistică se realizează prin preluarea datelor din diferite surse deja existente.

În general, probleme legate de timp şi loc apar în observarea statistică pentru unităţile statistice complexe, care includ şi elemente corespunzătoare unităţilor simple din care sînt formate.

Formulare şi instrucţiuni. Pentru a realiza o observare unitară este necesar ca formularele pe care se culeg datele statistice să fie tipizate şi întocmite pe baza programului observării, astfel încît să se poată da răspunsuri la toate caracteristicile.

La o observare statistică se pot folosi fie fişe individuale, fie liste. În fişele unei înregistrări statistice se trec datele cu privire la o singură unitate de observare, conform programului. Într-o listă se înscriu succesiv date pentru mai multe unităţi de observare statistică. Se pot folosi şi formulare combinate (fişe şi liste), în funcţie de situaţiile concrete existente în colectivitate. Completarea formularelor se face pe baza instrucţiunilor ţinînd seama de necesitatea de a obţine informaţii unitare pe tot teritoriul, pentru întreaga eclectivitate supusă observării.

Instrucţiunile de completare a formularelor trebuie să fie cît mai scurte, dar complete, asigurînd înţelegerea uniformă pentru toţi observatorii statistici. Ele pot să fie înscrise direct pe formular sau tipărite separat în broşuri. Instrucţiunile trebuie să precizeze nu numai aspectele metodologice ale culegerii datelor, ci şi problemele organizatorice.

Problemele organizatorice necesare desfăşurării normale a operaţiei de culegere a datelor. Observările statistice impun o pregătire minuţioasă atît din punct de vedere metodologic, cît şi organizatoric. Este necesar să se efectueze din timp o documentare amplă cu privire la experienţa acumulată cu prilejul organizării unor observări similare în trecut, în ţară sau pe plan mondial, în scopul asigurării obţinerii unei cantităţi cît mai mari de informaţie, de bună calitate, într-un timp cît mai scurt şi cu cheltuieli minime.

Între măsurile organizatorice necesare desfăşurării normale a unei observări statistice amintim : stabilirea listelor şi hărţilor necesare identificării tuturor unităţilor de la care urmează să se culeagă date statistice; efectuarea instructajului cu toţi observatorii, pentru ca să se asigure de la început o înţelegere unitară a tuturor aspectelor legate de observarea respectivă; operaţiile de tipărire şi difuzare a formularelor aprobate în prealabil de forurile competente.

Măsurile organizatorice sînt foarte variate de la o observare la alta, în funcţie de natura specifică a fenomenelor studiate, de sursele de informaţie utilizate, de volumul de date ce urmează să fie obţinut, de felul şi procedeul de înregistrare folosit. De aceea, planul unei observări trebuie

Scanned with OKEN Scanner

să se particularizeze pe fiecare cercetare în parte, în așa fel încît să se asigure datele necesare și suficiente caracterizării statistice a fenomenelor social-economice.

2.2. CLASIFICAREA OBSERVĂRILOR STATISTICE

Observările statistice sînt extrem de variate; uneori ele se combină în scopul realizării informațiilor necesare prelucrării statistice pentru obținerea unor indicatori sintetici, cu care se caracterizează colectivitățile statistice existente în mod obiectiv în cadrul societății, în fiecare etapă a dezvoltării lor.

Observările folosite în practica statistică se pot clasifica ținînd seama de următoarele criterii: modul de organizare; timpul observării; modul în care se caracterizează fenomenele; numărul unităților cuprinse în observare.

După modul de organizare, în practica statistică se pot întîlni observări *permanente* și *special organizate*.

Observările cu caracter permanent se realizează prin canalele sistemului informațional statistic, răspund direct funcțiilor statisticii de stat și sînt sintetizate în sistemul dărilor de seamă.

Pentru completarea informației statistice cu caracter permanent se organizează, după necesități, observări statistice speciale, de forma recensămintelor, anchetelor, monografiilor sau observărilor selective.

După timpul la care se referă datele, observările statistice pot fi *curente*, *periodice* și *unice*.

Observările statistice *curente* se bazează pe datele din sistemul informațional economic și urmăresc, în primul rînd, modul cum se realizează sarcinile de plan în toate compartimentele vieții sociale și economice. Ele furnizează date statistice necesare caracterizării continue a proceselor sociale și economice privite pe întreaga economie națională, pe ramurile ei, pe forme de proprietate și pe unități administrativ-teritoriale.

Observările *periodice* sînt cele care se efectuează la un anumit interval de timp și apar ca înregistrări statistice special organizate. De exemplu, recensămîntul populației are o periodicitate de aproximativ 10 ani, iar recensămîntul animalelor o periodicitate de un an.

Observările *unice* sînt cele care apar pentru consemnarea statistică a unui eveniment nerepetabil sau a unui eveniment repetabil întîmplător la intervale mari de timp.

După modul în care caracterizează fenomenele înregistrate se cunosc observări *statice și dinamice*.

Observările cu caracter static sînt acelea care surprind fenomenele la un moment dat. Pe baza acestor observări se pot caracteriza numai volumul și structura fenomenelor la un moment dat.

În observările cu caracter dinamic fenomenele sînt privite ca rezultat al unui proces social-economic, deci înregistrările se referă la o anumită perioadă. Pe baza unor astfel de observări se pot analiza nu numai variațiile produse în volumul și structura colectivităților studiate, dar și ten-

Scanned with OKEN Scanner

dinţele de dezvoltare pe o perioadă expirată, în vederea furnizării datelor pentru elaborarea diferitelor calcule de prognoză.

După numărul unităţilor cuprinse în observare se pot întîlni *observări totale* şi *observări parţiale*. Observările totale se caracterizează prin faptul că se culeg date de la toate unităţile colectivităţii. În ţara noastră ele se realizează prin sistemul dărilor de seamă şi prin recensăminte.

În observările parţiale se culeg informaţii numai pentru o parte a unităţilor colectivităţii. Ele servesc pentru a caracteriza statistic partea supusă înregistrării, sau pentru a caracteriza întregul ansamblu, deci pentru a înlocui o observare totală de mare amploare, care, din motive obiective, nu se poate realiza.

2.3. TIPURI DE OBSERVĂRI STATISTICE

În practica statistică se folosesc mai multe tipuri de observări statistice care au apărut şi s-au perfecţionat pe măsura dezvoltării statisticii şi a folosirii ei ca instrument de cunoaştere a realităţii social-economice. În procesul de perfecţionare a acestora un rol important l-a avut şi perfecţionarea instrumentelor de calcul care au permis o lărgire şi mai ales o adîncire a aplicabilităţii metodelor de observare statistică.

În practică se folosesc frecvent următoarele tipuri de observare statistică: sistemul dărilor de seamă; recensămîntul; observarea selectivă; observarea părţii principale; ancheta statistică; monografia.

2.3.1. Sistemul dărilor de seamă

Sistemul dărilor de seamă constituie principalul tip de observare statistică în economia socialistă. Existenţa sistemului informaţional economic şi social permite înregistrarea curentă a datelor statistice necesare pentru rezolvarea problemelor de bază ce interesează în mod permanent conducerea de partid şi de stat.

Dările de seamă reprezintă documente oficiale, întocmite de toate unităţile cuprinse în sistemul informaţional economic şi social. Informaţiile sînt înscrise pe formulare tipizate, care se completează şi se înaintează în termene strict stabilite.

Dările de seamă au ca obiectiv principal raportarea modului de realizare a sarcinilor de plan.

Fiind formulare tipizate, dările de seamă trebuie să cuprindă o serie de elemente de conţinut şi de formă, printre care cele mai importante sînt:

— elemente pe baza cărora se pot identifica unităţile care le completează (denumirea, adresa, subordonarea administrativă, funcţia şi numele celor care întocmesc şi răspund de exactitatea datelor înscrise într-o dare de seamă);

— elemente de identificare a dărilor de seamă şi de cuprindere a lor în sistemul de prelucrare şi transmitere a informaţiilor (titlul şi simbolul

Scanned with OKEN Scanner

dării de seamă, organul care a dat aprobarea, cu număr şi dată, perioada la care se referă, numărul exemplarelor şi către cine se transmit etc.).

Tipizarea formularelor de dări de seamă este impusă de necesitatea de a obţine de la toate unităţile şi în termen util a tuturor datelor statistice în vederea calculării indicatorilor centralizatori, ce sînt raportaţi în continuare către alt nivel de decizie. Pe baza datelor din dările de seamă se pot realiza funcţiile statisticii în economia socialistă şi se întocmesc anual balanţe statistice ale economiei naţionale, în care sînt analizate principalele proporţii şi corelaţii ale reproducţiei socialiste lărgite.

Dările de seamă apar ca observări statistice totale, permanente, care caracterizează fenomenele în continua lor desfăşurare.

Dările de seamă, deşi sînt completate periodic, reprezintă observări statistice curente, care pot apărea ca dări de seamă zilnice, decadale, lunare, trimestriale, semestriale şi anuale. Darea de seamă anuală este comună cu cea contabilă. Aceleaşi dări de seamă se pot întîlni în toate unităţile, apărînd ca dări de seamă comune ca, de exemplu, darea de seamă privind munca şi retribuirea după muncă, sau se întîlnesc numai într-un domeniu de activitate şi sînt considerate ca dări de seamă specifice ca, de exemplu, darea de seamă asupra realizării planului de mecanizare a lucrărilor agricole. Indiferent dacă sînt dări de seamă comune sau specifice, ele sînt observări totale, deoarece se completează la toate unităţile care au activitatea respectivă.

Pentru a avea o evidenţă clară asupra numărului dărilor de seamă, a periodicităţii întocmirii şi transmiterii lor pentru fiecare întreprindere şi ramură, se întocmeşte un *nomenclator* al dărilor de seamă. În nomenclator se înscriu toate dările de seamă care sînt obligatorii de completat pentru fiecare categorie de unităţi economice, organizaţie centrală sau instituţie social-culturală. Dările de seamă trebuie să fie înaintate beneficiarului de informaţie la termenele stabilite, completate în întregime şi de la toate unităţile care sînt cuprinse în sistemul informaţional economic.

Nomenclatoarele de dări de seamă se revizuiesc periodic în funcţie de necesităţile de informaţie statistică. Cu ajutorul lor se descoperă şi formularele care circulă neoficial şi nejustificat în cadrul unor departamente şi care trebuie eliminate.

Raţionalizarea şi simplificarea dărilor de seamă şi adaptarea lor continuă la cerinţele de conducere planificată a economiei naţionale rămîn o sarcină permanentă a organelor care răspund de buna funcţionare a sistemului informaţional naţional.

Cerinţele planificării economiei naţionale impun perfecţionarea sistemului de indicatori statistici, care să asigure nu numai o lărgire a ariei de investigaţie, dar mai ales o adîncire a procesului de cunoaştere statistică în domeniile în care ea se aplică în mod curent.

Pentru a putea furniza date cît mai complete şi cu un grad ridicat de operativitate, formularele de dări de seamă trebuie revizuite periodic, eliminîndu-se paralelismele în informaţia statistică, cît şi informaţiile devenite inutile. Pe această linie trebuie să se precizeze clar care sînt indicatorii strict necesari pentru fiecare etapă: decadă, lună, trimestru, an. Astfel, dările de seamă curente (zilnice, decadale, lunare, trimestriale) trebuie să cuprindă numai indicatori care să permită urmărirea sarcinilor

Scanned with OKEN Scanner

curente de plan şi luarea deciziilor în mod operativ, care să asigure mersul ritmic al realizării planului. Dările de seamă anuale trebuie să aibă un conţinut mai bogat, dîndu-se posibilitatea ca prin centralizarea datelor să se poată calcula, de exemplu, elementele necesare măsurării statistice a produsului social şi a venitului naţional, în vederea elaborării balanţei statistice a economiei naţionale.

Raţionalizarea şi simplificarea dărilor de seamă trebuie să aibă în vedere şi tehnica şi mijloacele cu care urmează să se prelucreze datele. Dotarea economiei naţionale cu echipamente moderne de calcul are ca rezultat organizarea fluxului informaţiilor economice în aşa fel, încît să asigure o creştere a calităţii şi operativităţii acestuia. Folosirea tehnicii moderne de prelucrare a datelor statistice trebuie să asigure un conţinut mai bogat sistemului de indicatori obţinuţi prin dările de seamă.

La fel, folosirea din ce în ce mai frecvent a informaţiilor obţinute prin cercetări selective permite restrîngerea numărului de indicatori obţinuţi în sistemul dărilor de seamă, deci vine în sprijinul raţionalizării şi simplificării acestuia.

2.3.2. Recensămîntul

Recensămîntul este cea mai veche formă de observare statistică. Ca observare statistică, recensămîntul se caracterizează prin faptul că este o înregistrare totală, special organizată, cu caracter periodic şi care surprinde fenomenele la un moment dat.

Recensămîntul este o înregistrare statică a fenomenelor; pentru analiza deplasărilor de structură de la o perioadă la alta se folosesc datele de la mai multe recensăminte succesive. De aceea este necesar ca programul observării să fie astfel întocmit încît să permită compararea datelor în dinamică.

De regulă, recensămîntul este o operaţie de mare amploare şi se caracterizează prin faptul că în etapa culegerii datelor sînt atrase şi persoane din afara aparatului statistic. De aceea, un loc important în planul observării pe bază de recensămînt îl ocupă atragerea şi instruirea persoanelor care vor deveni recenzori.

Orice recensămînt impune o serie de măsuri organizatorice legate de împărţirea teritoriului pe zone de recenzare, întocmirea de hărţi şi liste de unităţi de la care urmează să se culeagă date etc. La recensămintele de populaţie, în funcţie de densitatea populaţiei, se împarte teritoriul în sectoare de recensămînt, iar acestea în sectoare de înregistrare.

Pentru a putea realiza culegerea datelor în condiţii optime, este necesar să se facă recensăminte de probă prin care să se verifice toate punctele din programul observării. În instrucţiunile elaborate pentru un recensămînt trebuie să se dea cît mai multe exemple de situaţii concrete şi de modul în care ele trebuie să fie rezolvate.

Pentru a asigura un conţinut real indicatorilor obţinuţi pe baza unui recensămînt este necesar să se ia toate măsurile pentru a înţelege în mod unitar toate problemele legate de culegerea şi transmiterea datelor.

În cazul recensămîntului populaţiei, o primă centralizare a datelor se face chiar de către persoanele care au participat la observarea statis-

Scanned with OKEN Scanner

tică. În acest fel, prin centralizarea rezultatelor pentru un număr restrîns de indicatori pe întreaga ţară se obţin date preliminare, care urmează să fie confirmate după prelucrarea integrală a recensămîntului.

La recensămîntul populaţiei se poate combina observarea totală cu cea parţială. De exemplu, la recensămîntul din 1966, pe lîngă datele înregistrate cu privire la întreaga populaţie, s-a efectuat şi o observare selectivă în care s-au cules date referitoare la condiţiile de locuit ale populaţiei.

Recensămintele se folosesc şi pentru cunoaşterea fenomenelor economice, ca, de exemplu, recensămîntul mijloacelor materiale circulante, recensămîntul efectivului de animale etc. Aceste recensăminte se efectuează după aceleaşi principii ca şi cele de populaţie, dar au periodicităţi mult mai reduse (anual sau de două ori pe an).

2.3.3. Observarea selectivă

Este o înregistrare parţială, care se organizează ori de cîte ori nu se poate efectua o observare totală, sau cînd este justificată prin operativitate, economicitate şi condiţii specifice de realizare. Rezultă că observarea selectivă trebuie să fie astfel organizată încît să se poată înlocui o observare totală. Datele culese prin această metodă trebuie să fie reprezentative pentru întreaga colectivitate.

Aplicarea metodei selective impune rezolvarea unor probleme legate de alegerea eşantionului şi de calculul erorilor de estimare a indicatorilor pe întreaga colectivitate, folosind principiile teoriei probabilităţilor. Acest tip de observare se va trata într-un capitol separat intitulat „Cercetarea selectivă".

2.3.4. Observarea părţii principale

Observarea părţii principale se foloseşte atunci cînd se studiază o colectivitate care prezintă variaţii calitative substanţiale de la o grupă la alta, astfel încît unele grupe au o influenţă hotărîtoare în formarea indicatorilor pe întreaga colectivitate, iar altele au o influenţă nesemnificativă. În acest caz este suficient să se supună observării numai partea principală a colectivităţii şi, cu anumite rezerve, să se caracterizezeîntregul an samblu. De exemplu, dacă se măsoară rezervele de creştere a producţiei în industria R. S. România, este suficient să se ia în observare întreprinderile hotărîtoare ale industriei pentru a aprecia volumul resurselor ce pot fi atrase şi valorificat în procesul de producţie.

În acest caz, informaţiile ce vor fi culese nu au caracter reprezentativ, dar pot da o informaţie orientativă asupra tendinţelor ce se manifestă la nivelul întregului.

2.3.5. Anchetele statistice

Anchetele statistice sînt observări parţiale, special organizate, care, de regulă, se bazează pe o completare benevolă a chestionarelor. Deci şi în acest caz nu se poate pune problema asigurării gradului de reprezenta-

Scanned with OKEN Scanner

tivitate a părţii observate faţă de colectivitatea totală. Pe baza datelor unei anchete statistice se apreciază care sînt tendinţele generale care se manifestă în cadrul colectivităţii. Un exemplu de anchetă statistică este cea efectuată cu prilejul organizării unei expoziţii de mărfuri pentru contractare. În fiecare stand de mărfuri există chestionare, care sînt completate benevol de vizitatori, în care se indică mărfurile care ar dori să fie contractate de comerţul cu amănuntul. Este clar că unităţile comerţului cu amănuntul nu pot să structureze fondul de mărfuri numai pe baza prelucrării datelor din aceste chestionare, dar la contractări vor avea în vedere şi frecvenţa cu care au fost solicitate mărfurile respective. La studiul cererii de mărfuri metoda de bază este aceea a observării selective, în care toate categoriile populaţiei sînt reprezentate şi deci, cu un anumit grad de siguranţă, se pot estima şi indicatorii pe întreaga colectivitate.

2.3.6. Monografia

Monografia se foloseşte frecvent în studiile statistice pentru caracterizarea unei unităţi statistice complexe. De regulă, prin studiile monografice se urmăreşte cunoaşterea elementelor noi apărute în economia unei ţări şi care vor deveni fenomene de masă.

Monografia statistică presupune o caracterizare multilaterală a unei unităţi complexe şi ca atare necesită folosirea unei echipe de cercetători. Caracteristic monografiilor este faptul că întreaga muncă de cercetare, de la culegerea la interpretarea rezultatelor se face în mod unitar de către aceeaşi echipă de cercetători. Cu cît unitatea este mai complexă, cu atît este necesar să fie folosite echipe mixte de cercetători din toate domeniile care interesează. De pildă, se pot întocmi monografii de întreprinderi industriale şi agricole, monografii de sate, oraşe sau judeţe şi monografii pe probleme (studiul monografic al productivităţii muncii în industria unei ţări, într-o ramură etc.).

2.4. ERORILE OBSERVĂRII STATISTICE

Principiul de bază al efectuării unei observări statistice este acela al asigurării autenticităţii datelor. Prin autenticitatea datelor se înţelege corespondenţa deplină între datele înregistrate şi mărimea obiectivă a fenomenului. Buna organizare a unei observări statistice trebuie să se reflecte în obţinerea unor date exacte şi în volum complet.

Practica statistică a dovedit că în mod curent se pot produce *erori de înregistrare.* Aceste erori de înregistrare sînt cu atît mai numeroase cu cît cercetarea este de mai mare amploare şi cuprinde un program amplu de observare.

Erorile de observare reprezintă abateri între datele înregistrate şi mărimea concretă, reală a caracteristicilor cuprinse în programul observării. Aceste erori pot să aibă caracter întîmplător sau sistematic.

Erorile de observare întîmplătoare se produc, de regulă, în ambele sensuri şi sînt făcute nepremeditat, fie din neînţelegerea corectă a întrebării, fie din lipsă de memorie. Cu cît observarea se referă la un număr

Scanned with OKEN Scanner

mai mare de unităţi, cu atît posibilităţile de compensare a acestor erori sînt mai mari şi ca atare ele vor influenţa în mai mică măsură rezultatele de ansamblu.

Erorile de observare sistematice se produc, de regulă, într-un singur sens şi influenţează asupra indicatorilor de ansamblu.

Pe lîngă erorile de observare comune oricărei înregistrări se întîlnesc şi erori de reprezentativitate specifice cercetării selective şi care se interpretează în funcţie de probabilitatea cu care se garantează rezultatele selecţiei pentru estimarea indicatorilor din colectivitatea totală (vezi capitolul „Cercetarea selectivă").

2.5. CONTROLUL DATELOR STATISTICE

Controlul are drept scop de a descoperi eventualele erori de observare şi de a asigura autenticitatea datelor culese. Controlul datelor statistice este deosebit de important, pentru că în procesul de prelucrare datele obţinute prin observare îşi pierd individualitatea şi erorile se regăsesc în valoarea indicatorilor derivaţi.

Verificarea datelor statistice începe printr-un *control de volum* prin care se constată dacă s-au completat toate formularele şi toate rubricile, apoi se trece la efectuarea controlului aritmetic şi a controlului logic.

Controlul aritmetic presupune efectuarea unor operaţii aritmetice pentru verificarea indicatorilor numerici care figurează în acelaşi formular statistic. De exemplu, într-o dare de seamă asupra îndeplinirii planului de muncă şi retribuire după muncă, retribuţia medie se obţine ca raport între fondul de retribuire şi numărul de muncitori. Aceşti trei indicatori se găsesc în darea de seamă şi se pot verifica relaţiile cantitative dintre ei.

Controlul logic se efectuează prin compararea valorilor care se găsesc în anumite relaţii de interdependenţă. În exemplul precedent retribuţia medie pe total trebuie să se încadreze ca valoare între nivelurile retribuţiei medii pe fiecare categorie de personal şi se apropie de retribuţia care are ponderea cea mai mare. Dacă valoarea pe total, înscrisă în formular, depăşeşte limitele extreme ale retribuţiilor primite, înseamnă că s-a strecurat o greşeală care trebuie să fie înlăturată. Tot printr-un control logic se pot elimina eventualele erori, dacă se compară în timp şi spaţiu unităţile care au condiţii asemănătoare şi prezintă variaţii foarte mari ale indicatorilor înscrişi în formularele de observare.

Controlul logic se mai poate efectua şi pe baza unor întrebări de control înscrise în acest scop în formularele de înregistrare pentru anumite caracteristici.

Controlul datelor statistice trebuie să se facă şi pentru înregistrările pe bază de documente, în sensul analizelor periodice a modului în care se ţin în fiecare unitate documentele folosite pentru a completa dările de seamă.

Folosind combinat aceste forme de control se poate verifica gradul de exactitate al datelor obţinute într-o observare statistică. Controlul statistic face legătura dintre observarea şi prelucrarea datelor în procesul de cercetare statistică.

Scanned with OKEN Scanner

CAPITOLUL III

PRELUCRAREA DATELOR STATISTICE

În sensul cel mai larg, prin prelucrarea statistică se înțelege etapa cercetării statistice în care se trece de la datele (informațiile) individuale obținute în etapa observării pentru fiecare unitate a colectivității, la indicatori care caracterizează colectivitatea în întregul ei, sub formă de indicatori totalizatori absoluți sau indicatori derivați. De regulă, acești indicatori sînt calculați ca valori sintetice obținute prin aplicarea unor abstractizări și comparări succesive, prin care se elimină ceea ce este întîmplător și neesențial în apariția și dezvoltarea fenomenelor studiate. Prin prelucrare se obține sistemul de indicatori statistici, ca parte integrantă a sistemului informațional național, necesar pentru fundamentarea deciziilor la toate nivelele de conducere.

Măsurile luate de conducerea de partid și de stat au în vedere continua perfecționare a sistemului informațional, care impune tocmai obținerea unei informații statistice complete, care să poată fi folosită la fundamentarea deciziilor necesare unei conduceri științifice a întregii economii. În articolul 1 din paragraful 2 al Hotărîrii C.C. al P.C.R. cu privire la perfecționarea sistemului informațional economico-social, introducerea sistemelor de conducere cu mijloace de prelucrare automată a datelor și dotarea economiei naționale cu tehnică de calcul în perioada 1971—1980, se arată: „se va îmbunătăți sistemul de indicatori economico-sociali în vederea caracterizării multilaterale și obiective a fenomenelor și proceselor economice, a modului de îndeplinire a planului; se va urmări, de asemenea, asigurarea unității metodologice și de conținut a indicatorilor de plan, statistici și financiari, coordonarea întregului sistem de indicatori" [1].

Sarcini deosebite, pe linia aplicării măsurilor de perfecționare a indicatorilor economico-financiari, revin statisticii în urma Plenarei Comitetului Central al Partidului Comunist Român din 22—23 martie 1978, cuprinse în Hotărîrea C.C. al Partidului Comunist Român cu privire la perfecționarea conducerii și planificării economico-financiare [2], în special prin introducerea ca indicator de bază a valorii producției nete.

Tehnica modernă de prelucrare a informației impune o reconsiderare atît a modului de utilizare a surselor de informație, cît și a metodei de prelucrare și analiză statistică a fenomenelor studiate. În prelucrarea mecanizată și automatizată a datelor statistice crește considerabil rolul analizei. Conceperea cercetării statistice ca o cercetare sistemică necesită o cunoaștere teoretică și practică cît mai adîncită a structurii interne a fenomenelor și a tuturor trăsăturilor cu caracter stabil, astfel încît să se realizeze o legătură permanentă între datele prelucrate și cele care urmează să fie culese în etapa următoare. În același timp, folosirea mijloacelor moderne de calcul permite aplicarea celor mai variate metode și procedee statistice pentru o caracterizare cantitativă cît mai analitică a tipurilor cali-

[1] Scînteia, 22 aprilie 1972.
[2] Scînteia, 28 martie 1978.

Scanned with OKEN Scanner

tative existente, a mutaţiilor produse în structura internă a fenomenelor pe plan teritorial sau în dinamică, precum şi a altor aspecte ce nu pot fi studiate decît prin mijloace statistice.

Conţinutul prelucrării unei observări statistice trebuie să se refere la: centralizarea materialului observării; gruparea unităţilor colectivităţii; obţinerea sistemului de indicatori ce caracterizează fiecare grupă şi întreaga colectivitate; prezentarea rezultatelor prelucrării sub formă de serii, tabele şi grafice.

Din motive didactice, în acest capitol va fi prezentată numai faza de centralizare şi sistematizare a materialului obţinut în urma observării statistice, deci numai obţinerea sistemului de indicatori totalizatori, generali şi pe grupe.

Complexitatea acestor probleme impune stabilirea unui plan în care să fie precizate atît aspectele metodologice, cît şi cele organizatorice.

3.1. PLANUL PRELUCRĂRII STATISTICE

Planul prelucrării materialului unei observări statistice trebuie să cuprindă: programul prelucrării; metodele şi procedeele de calcul statistic ce se vor aplica pentru obţinerea sistemului de indicatori; formele de prezentare a rezultatelor prelucrării; aspectele organizatorice ale prelucrării.

Programul prelucrării constituie problema esenţială a conţinutului prelucrării şi constă în enumerarea caracteristicilor primare şi derivate care se folosesc pentru calcularea indicatorilor totalizatori şi derivaţi. Dacă prelucrarea se face pe baza unei cercetări statistice, special organizate, atunci planul şi respectiv programul prelucrării se întocmesc înainte de declanşarea cercetării, odată cu planul general al prelucrării statistice. Dacă prelucrarea se face folosind informaţia deja existentă în cadrul sistemului informaţional economic, atunci planul prelucrării trebuie să se elaboreze în funcţie numai de caracteristicile pentru care se culeg în mod curent date statistice. Dacă aceste date nu sînt suficiente, atunci este necesar să se suplimenteze informaţiile existente printr-o observare totală sau selectivă, după caz.

Metodele şi procedeele de calcul statistic trebuie să se aleagă pentru fiecare prelucrare în parte, în funcţie de scopul cercetării, de natura specifică a fenomenelor şi a informaţiei de care se dispune.

Aceasta înseamnă că metodele şi procedeele de prelucrare trebuie să fie diferenţiate în măsura în care fenomenele se modifică de la o etapă la alta. Scopul principal al aplicării metodelor de prelucrare statistică îl constituie redarea într-o imagine cît mai sintetică şi obiectivă a fenomenelor pe care le studiază şi care, într-un sens mai larg, ar corespunde noţiunii de modelare statistică. În această etapă este necesar, deci, să se elaboreze

Scanned with OKEN Scanner

fie modelele de calcul și analiză statistică, fie să se selecteze modelele deja elaborate, care se pot adapta la specificul datelor primare culese.

Formele de prezentare a rezultatelor prelucrării statistice sînt : seriile statistice, tabelele statistice, graficele.

În planul prelucrării statistice trebuie să se precizeze care sînt machetele tabelelor de prelucrare și de prezentare, nomenclatorul caracteristicilor după care se face centralizarea datelor, formele de grafice ce se vor utiliza atît pentru interpretarea datelor în vederea aplicării calculelor statistice, cît și pentru prezentarea rezultatelor obținute.

Problemele organizatorice ale prelucrării trebuie să se refere la locul și timpul cînd se face prelucrarea și transmiterea rezultatelor prelucrării. Dacă cercetarea este de o mai mare amploare și se fac și prelucrări parțiale, atunci se vor preciza și termenele și gradul de prelucrare, atît pentru fiecare fază a prelucrării, cît și pentru întreaga lucrare.

3.2. TEHNICI DE PRELUCRARE

Prelucrarea statistică este, în general, o operație de mare amploare care se face, de regulă, de instituții specializate. Complexitatea problemelor metodologice și organizatorice pe care le implică, mai ales după introducerea prelucrării automatizate a datelor statistice, au dus la introducerea în acest domeniu a unor discipline de predare în învățămînt. De aceea, în acest manual nu se vor face decît referiri la modul de organizare a prelucrării datelor statistice.

Dezvoltarea tehnicii de calcul și dotarea economiei naționale cu echipamente de calcul au ca rezultat creșterea continuă a ponderii prelucrării mecanografice și electronice a datelor statisticii curente. Aceasta se reflectă în creșterea considerabilă a cantității de informație statistică și a operativității obținerii ei.

După tehnica de calcul folosită, prelucrarea poate fi efectuată manual, mecanografic și cu ajutorul calculatoarelor electronice. Vor fi prezentate, din considerentele arătate mai sus, numai cîteva elemente de prelucrare manuală a informației statistice.

Prelucrarea manuală ocupă un loc din ce în ce mai restrîns și se folosește în special în cercetări de volum mai mic și, de regulă, se completează cu efectuarea unor calcule la mașinile de birou. Prelucrarea manuală se utilizează atît pentru centralizarea unităților, cît și pentru centralizarea valorilor caracteristicilor din programul observării și se realizează fie prin procedeul fișelor, fie prin procedeul tabelării manuale.

Procedeul fișelor constă în întocmirea unei fișe pentru fiecare unitate de observare în care se trec — de obicei codificat — toate caracteristicile din programul observării. Cu ajutorul fișelor se pot face grupări simple sau grupări combinate. În general, procedeul fișelor se folosește cînd există

Scanned with OKEN Scanner

un număr mai mic de unităţi observate şi se fac grupări în funcţie de mai multe caracteristici statistice. Acest procedeu prezintă, deci, avantajul că se pot obţine cîte variante de grupare sînt necesare şi dezavantajul că nu se pot elimina eventualele erori, decît reluînd toate operaţiile de prelucrare.

Procedeul tabelării manuale se foloseşte în cazul cînd volumul colectivităţii este mare şi nu este posibilă prelucrarea cu ajutorul echipamentelor mecanografice sau electronice; de obicei se obţin grupări simple prin centralizarea celorlalte caracteristici cu care caracteristica de grupare se găseşte în interdependenţă.

În cazul tabelării manuale, în capetele coloanelor tabelului se trec caracteristicile (inclusiv caracteristicile de grupare) iar pe fiecare rînd se înscrie cîte o unitate de observare cu informaţiile concrete corespunzătoare acestor caracteristici.

Acest procedeu prezintă avantajul că se pot depista cu uşurinţă erorile şi dezavantajul că dacă se schimbă structura grupelor sau caracteristica de grupare, trebuie să se facă din nou tabelarea.

3.3. CENTRALIZAREA DATELOR STATISTICE

După cum s-a mai arătat, datele statistice obţinute prin observare sînt extrem de variate, provenind de la un număr foarte mare de unităţi de înregistrare purtătoare a unor caracteristici cu un grad mare de împrăştiere. În aceste condiţii este greu de reţinut cum se face repartiţia acestor valori individuale, care este legătura dintre două sau mai multe caracteristici înregistrate, care sînt părţile componente ale colectivităţii. Pentru a realiza o primă imagine a fenomenelor înregistrate este necesar să se facă o sistematizare a acestui material „brut".

Această sistematizare se realizează prin operaţia de centralizare a materialului unei observări statistice şi presupune atît strîngerea la locul de prelucrare a tuturor informaţiilor (formularelor), cît şi operaţia de obţinere a indicatorilor totalizatori.

Centralizarea datelor statistice poate fi simplă şi pe grupe.

3.3.1. Centralizarea simplă

Centralizarea simplă presupune numai obţinerea indicatorilor totalizatori la nivelul întregului ansamblu de unităţi care au fost supuse observării. Ea presupune stabilirea volumului unităţilor simple şi complexe care formează colectivitatea statistică, precum şi valoarea totalizată a caracteristicilor care au fost incluse în programul observării şi care sînt exprimate în unităţi de măsură însumabile direct. În cazul cînd nu se

Scanned with OKEN Scanner

poate face însumarea directă a datelor individuale, se folosesc anumiți coeficienți de echivalență care permit apoi centralizarea datelor. În sfîrșit, dacă s-au folosit unități complexe de observare, la care s-au înregistrat și caracteristici exprimate sub formă de mărimi relative sau medii, atunci și la nivelul colectivității ele trebuie centralizate după aceeași metodă după care au fost calculate.

Centralizarea simplă se folosește, de regulă, pentru obținerea unor informații generale, cu mult înainte de a se finaliza întreaga operație de prelucrare. Este cazul unor observări statistice de mare amploare pentru care durata de prelucrare este foarte mare, cum este, de exemplu, recensămîntul general al populației.

3.3.2. Centralizarea pe grupe

Centralizarea pe grupe presupune aplicarea metodei grupării în urma căreia se obțin indicatori totalizatori parțiali, pe fiecare grupă, iar pe baza lor se obțin apoi indicatori totalizatori generali, pe întreaga colectivitate. Deci, centralizarea pe grupe include și pe cea simplă.

Centralizarea pe grupe este operația cea mai frecvent întîlnită în practica statistică. Ea permite sistematizarea și pregătirea materialului obținut din observare pentru aplicarea metodelor și procedeelor de calcul și analiză statistică. Cu ajutorul ei se realizează nu numai o cunoaștere a întregului ansamblu, dar și a elementelor sale componente. Pentru realizarea acestui scop este necesar ca gruparea să nu fie considerată ca o operație strict tehnică, ci ca o metodă de bază, în care accentul trebuie să cadă pe implicațiile calitative ale aplicării ei și care se vor reflecta în calitatea întregului proces de cunoaștere statistică.

3.4. METODA GRUPĂRII — METODĂ DE BAZĂ ÎN PRELUCRAREA DATELOR STATISTICE

Gruparea statistică reprezintă prima sistematizare a unor date individuale, adesea foarte numeroase, care au fost obținute în urma unei observări statistice. Sistematizarea datelor statistice se face în scopul identificării fenomenelor de care aparțin aceleași esențe social-economice și prezintă, deci, o serie de proprietăți comune.

Separînd unitățile în grupe omogene și comparînd datele pentru aceleași caracteristici, distribuite în condiții variabile de timp și spațiu, se pot desprinde tendințele de repetabilitate a fenomenelor și se pot identifica și ierarhiza factorii care determină variația lor.

După cum s-a mai arătat, prin prelucrarea statistică datele observării de masă își pierd individualitatea, iar nivelul de dezvoltare sau forma de manifestare a fenomenelor se regăsește în valoarea indicatorilor totaliza-

Scanned with OKEN Scanner

tori sau sintetici care se obţin prin aplicarea metodelor de calcul statistic. Pentru ca aceşti indicatori să aibă conţinut real este necesar ca ei să fie obţinuţi din date primare omogene între ele; omogenizarea datelor statistice se realizează cu ajutorul metodei grupării.

Prin *grupare statistică* se înţelege împărţirea unităţilor în grupe omogene în funcţie de variaţia uneia sau a mai multor caracteristici. Caracteristica în funcţie de care se face separarea unităţilor se mai numeşte şi factor de grupare. Pentru a realiza o grupare în sens statistic, este necesar să se aleagă acele caracteristici esenţiale care au caracter stabil, determină în mod hotărîtor fenomenele care se studiază şi în funcţie de variaţia lor se pot obţine grupe omogene.

Prin *grupă omogenă* din punct de vedere statistic se înţelege clasa de unităţi în interiorul căreia variaţia caracteristicii este minimă. Această variaţie din interiorul grupei este interpretată statistic ca fiind rezultatul influenţei factorilor întîmplători şi care, la nivelul întregului ansamblu, se compensează reciproc.

Fenomenele studiate de statistică depind, de regulă, de mai mulţi factori esenţiali. Aceasta înseamnă că, de cele mai multe ori, pentru a obţine grupe omogene trebuie să se folosească mai multe caracteristici de grupare, ţinînd seama de interdependenţa obiectivă dintre ele.

O contribuţie importantă la fundamentarea metodei grupării datelor statistice a adus-o V.I. Lenin. Pentru caracterizarea fenomenelor şi proceselor din economia capitalistă, Lenin a studiat un bogat material informativ din statisticile oficiale din vremea sa şi a arătat, pe baza unor argumente ştiinţifice, că dacă gruparea datelor statistice nu este corectă, ea poate duce la estomparea deosebirilor calitative existente în cadrul colectivităţilor studiate şi face imposibilă cunoaşterea obiectivă a aspectelor esenţiale specifice fiecărei etape a dezvoltării lor. El a arătat, de asemenea, că o serie de date preţioase obţinute prin recensămintele efectuate în diferite ţări capitaliste nu pot fi folosite corect, deoarece: folosesc caracteristici de grupare neesenţiale; folosesc un număr prea mare de grupe ce duc la fărîmiţarea fenomenului; folosesc un număr prea mic de grupe în scopul ascunderii deosebirilor dintre tipurile calitative care s-au conturat în mod obiectiv în interiorul lor; alegerea mărimii intervalelor de grupare nu este corespunzătoare cu gradul de variaţie a factorilor de grupare. Regrupînd datele statistice din vremea sa şi comparînd noile rezultate obţinute prin folosirea unor grupări ştiinţifice fundamentate, el a reuşit în opera „Dezvoltarea capitalismului în Rusia" să demonstreze că în mod obiectiv şi în Rusia ţaristă se manifesta tendinţa de concentrare şi centralizare a capitalului nu numai în industrie, ci şi în agricultură.

Ţinînd seama de observaţiile critice şi de indicaţiile date de Lenin în legătură cu principiile de grupare a datelor statistice, aplicarea metodei grupărilor în statistica socialistă presupune rezolvarea următoarelor probleme: alegerea şi folosirea caracteristicilor de grupare; stabilirea numărului de grupe în care urmează să se împartă unităţile colectivităţii; alegerea mărimii intervalului de grupare; precizarea scopului pentru care se face gruparea.

Scanned with OKEN Scanner

3.4.1. Alegerea și folosirea caracteristicilor de grupare

Programul observării statistice poate să cuprindă atît caracteristici esențiale, determinante pentru variația altor caracteristici, cît și unele caracteristici care au caracter secundar sau cu rol de control. Ele sînt folosite pentru caracterizarea multilaterală a fenomenelor studiate în raport cu scopul cercetării.

În funcție de scopul grupării, caracteristicile statistice pot să-și schimbe gradul de esențialitate. De exemplu, în programul unei observări statistice pentru ramura industriei figurează, printre altele, următoarele caracteristici: mărimea fondurilor fixe; mărimea producției globale; numărul muncitorilor; productivitatea muncii; costul unitar al producției. Dacă interesează gradul de mărime al întreprinderilor, atunci se va lua în considerare variația primelor trei caracteristici. Productivitatea muncii ca și costul producției, deși sînt caracteristici esențiale pentru fiecare întreprindere, totuși ele sînt auxiliare pentru caracterizarea gradului de mărime al acestora. Dacă, însă, s-ar analiza căile de reducere a costurilor producției în industrie, atunci productivitatea muncii și costul producției ar fi caracteristici la fel de esențiale ca și gradul de înzestrare tehnică a muncii — calculat ca raport între valoarea fondurilor fixe și numărul de muncitori —, precum și mărimea și structura producției. Înseamnă că alegerea caracteristicilor de grupare trebuie să se facă astfel, încît ele să corespundă scopului analizei statistice și, totodată, să împartă unitățile colectivității în grupe omogene.

3.4.2. Felurile grupărilor statistice

În practică grupările statistice se folosesc în toate domeniile vieții sociale și economice. Fiind extrem de numeroase și variate, ele trebuie să fie analizate după mai multe criterii.

După numărul caracteristicilor folosite pot fi grupări simple și grupări combinate.

Grupările simple sînt acelea în care gruparea unităților se face în funcție de variația unei singure caracteristici și se pot centraliza atît numărul unităților, cît și valorile unor caracteristici cu care variabila de grupare se găsește în interdependență. Grupările simple se pot efectua pentru toate caracteristicile din programul observării sau numai pentru acelea care interesează mai departe procesul de prelucrare.

Grupările combinate sînt acelea în care se folosesc împreună mai multe caracteristici pentru formarea grupelor. Se alege o primă caracteristică de grupare și, în funcție de variația ei, se repartizează unitățile colectivității; în continuare, fiecare grupă se separă pe subgrupe după variațiile celei de-a doua caracteristici de grupare; la rîndul său, fiecare subgrupă se separă după variația celei de-a treia caracteristici de grupare ș.a.m.d. În funcție de creșterea numărului caracteristicilor de grupare se obțin grupe din ce în ce mai omogene. Pentru a nu fărîmița colectivitatea în grupe prea numeroase se recomandă să nu se depășească 3—4 caracteristici de grupare.

Scanned with OKEN Scanner

După conținutul caracteristicii de grupare se pot clasifica în grupări cronologice, grupări teritoriale și grupări după o caracteristică atributivă.

Grupările cronologice și teritoriale sînt determinate de condițiile obiective de timp și loc în care se produc fenomenele. Dar nu orice sistematizare a informației statistice după variația fenomenelor în timp și spațiu presupune o grupare în sensul statistic precizat mai înainte. Pentru ca separarea unităților și a valorilor caracteristicilor înregistrate să formeze o grupare cronologică trebuie ca variația de timp să constituie o caracteristică determinantă pentru datele care se grupează. De exemplu, pentru populație vîrsta (respectiv anul nașterii) este o caracteristică esențială, la fel în statistica sanitară perioada de spitalizare. Același lucru este valabil și pentru caracteristica teritorială. Dacă variația teritorială este semnificativă pentru formarea fenomenelor studiate, atunci folosirea ei conduce la o grupare tipologică. De exemplu, pentru producția agricolă, separarea datelor statistice pe zone climaterice are ca rezultat o grupare tipologică a datelor.

De subliniat că atît variația de timp, cît și cea de spațiu trebuie să fie studiată prin metode statistice. De aceea, centralizarea pe grupe a datelor statistice trebuie să fie făcută și după variația acestor caracteristici, chiar dacă n-au conținut de grupare tipologică. În astfel de grupări, pentru celelalte caracteristici care interesează, se face o grupare condiționată de timpul sau de spațiul în interiorul cărora se găsesc aceste caracteristici.

Grupările după o caracteristică atributivă se întîlnesc frecvent în cercetarea statistică fie ca grupări simple, fie ca grupări combinate, ajutînd la caracterizarea independentă sau interdependentă a fenomenelor.

Caracteristicile atributive definesc colectivitatea statistică și separă calitativ fenomenele între ele. După forma de exprimare ele pot fi : caracteristici statistice exprimate prin cuvinte și caracteristici statistice exprimate numeric.

Grupările după caracteristici atributive exprimate prin cuvinte se întîlnesc în statistica social-economică, de regulă, sub denumirea de *clasificări* și sînt grupări cu caracter oficial. Clasificările se cuprind în nomenclatoare ce se revizuiesc periodic pentru a comasa unele grupe care își pierd din importanță, sau a dezvolta altele care, în procesul de diferențiere a fenomenelor încep să devină eterogene. Exemple : clasificarea ramurilor economiei naționale, folosită pentru gruparea tuturor indicatorilor statistici care permit caracterizarea procesului reproducției socialiste lărgite, gruparea personalului muncitor după profesii și meserii.

Grupările după o caracteristică atributivă exprimată prin cuvinte sînt cele mai omogene, deoarece în aceeași grupă se includ unități care, din punctul de vedere al caracteristicii de grupare, nu prezintă variație. De exemplu, la gruparea pe meserii, în grupa de strungari se vor include numai cei care au obținut calificarea în acest domeniu. Desigur că și această grupă va prezenta o variație statistică, dar luată în funcție de celelalte caracteristici, ca de pildă : vîrsta, vechimea în meserie, gradul de îndemînare etc.

Pentru a înțelege tehnica de grupare a datelor după o caracteristică numerică este necesar să se facă unele precizări în legătură cu următoarele

Scanned with OKEN Scanner

noţiuni : variantă, variabilă, variaţie, amplitudinea variaţiei, frecvenţă sau pondere.

Varianta este nivelul sau valoarea prin care se manifestă o caracteristică statistică, înregistrată într-o observare.

Variabila este caracteristica statistică care prezintă proprietatea de a-şi modifica nivelul de la o unitate statistică la alta.

Variaţia este proprietatea caracteristicilor statistice de a-şi modifica nivelul de dezvoltare într-un interval de valori, în condiţii date de timp şi spaţiu, în funcţie de gradul de influenţă a factorilor determinanţi.

Amplitudinea variaţiei reprezintă cîmpul de împrăştiere a variantelor într-o înregistrare statistică. Ea se poate măsura ca diferenţă între variantele maxime şi variantele minime înregistrate în aceeaşi observare statistică.

Frecvenţa sau *ponderea* este numărul de unităţi la care se întîlneşte acelaşi nivel de dezvoltare sau formă de manifestare. Frecvenţa se poate exprima fie prin mărimi absolute, fie prin mărimi relative de structură [1].

În cazul cînd pentru grupare se folosesc caracteristici exprimate numeric este necesar să se aleagă numărul de grupe şi dacă variaţia este mare să se stabilească şi mărimea intervalului de grupare.

După modul de variaţie, pentru caracteristicile exprimate numeric, în practica statistică se întîlnesc trei tipuri de grupări : grupări pe variante; grupări pe intervale egale de variaţie; grupări pe intervale neegale de variaţie.

Grupările pe variante se folosesc atunci cînd numărul variantelor este redus şi deci centralizarea datelor se poate face pentru fiecare variantă în parte. De exemplu : producţia construcţiilor de locuinţe se poate grupa după numărul de camere; în cercetarea socio-demografică familiile se pot grupa după numărul de persoane din care sînt compuse sau după numărul persoanelor active.

Grupările pe intervale egale de variaţie se folosesc cînd gradul de variaţie al caracteristicii permite alegerea unei mărimi egale a intervalelor, astfel încît numărul grupelor să nu modifice forma ei de variaţie.

Grupările pe intervale neegale se folosesc fie pentru cazul unui grad foarte marc de variaţie, fie gentru grupările tipologice unde diferenţierea pe tipuri calitative cere intervale neegale.

3.4.3. Alegerea numărului de grupe şi stabilirea mărimii intervalului de grupare pentru caracteristicile exprimate numeric

Alegerea numărului de grupe şi a intervalului de grupare se face ţinînd seama de scopul pentru care se foloseşte metoda grupării. Dacă datele trebuie să fie supuse în continuare prelucrării statistice în vederea calculării indicatorilor derivaţi, atunci trebuie să se stabilească un număr mai mare de grupe şi pe cît posibil cu intervale egale de variaţie de la o grupă la alta. Deci, în acest caz gruparea are rol de sistematizare a datelor observării de masă. Dacă gruparea se face ca mijloc de analiză în vederea sta-

[1] Vezi mărimile relative de structură în capitolul : „Mărimi relative".

Scanned with OKEN Scanner

bilirii structurii pe tipuri calitative şi a mutaţiilor intervenite în structura colectivităţii, atunci se pot folosi un număr mai mic de grupe şi intervale neegale de variaţie de la o grupă la alta.

Indiferent de scopul grupării, alegerea numărului de grupe şi a intervalelor de grupare trebuie să se facă după ce s-au precizat relaţiile obiective de cauzalitate ce se manifestă în interiorul colectivităţii studiate şi în aşa fel să nu fărîmiţeze fenomenele şi nici să ascundă diferenţierea pe tipurile calitative conturate în mod obiectiv în interiorul colectivităţii statistice.

Pentru cazul în care gruparea se face în vederea sistematizării datelor se ia următorul exemplu : într-o echipă de muncitori s-au înregistrat următoarele valori ale producţiei individuale : 125; 128; 130; 131; 142; 135; 136; 142; 136; 143; 125; 123; 135; 123; 132; 143; 133; 132; 135; 122; 128; 135; 124; 131; 134; 125.

Se poate observa că pentru acest număr mic de date statistice este dificil de format o imagine uşor de reţinut pentru a caracteriza nivelul producţiei în echipa respectivă. Pentru aceasta este necesar să se ordoneze datele, să se obţină frecvenţele de apariţie a diferitelor variante. În acest scop, se porneşte de la amplitudinea variaţiei şi de la numărul unităţilor observate.

Dacă se notează caracteristica statistică după care se grupează cu (x), ea poate lua valori între limita minimă (x_{min}) şi limita maximă (x_{max}).

Amplitudinea variaţiei $A = x_{max} - x_{min}$;

$$x_{min} = 122;\ x_{max} = 143;\ \text{deci}\ A = 143 - 122 = 21.$$

Numărul de grupe (r) şi mărimea intervalului de grupare (K) se aleg în aşa fel încît să se cuprindă toate valorile individuale. Ţinînd seama şi de numărul mic al unităţilor observate şi de amplitudinea variaţiei se poate aprecia că sînt suficiente 5 grupe, deci : $K = \frac{A}{r} = \frac{21}{5} = 4{,}20 \approx 5$.

Dacă din raportul obţinut între mărimea amplitudiinii variaţiei şi numărul grupelor se obţine un număr zecimal, se rotunjeşte totdeauna în plus, astfel încît mărimea intervalelor de grupare să dea posibilitatea perceperii mai uşoare şi rapide a variaţiei caracteristicii de grupare şi să includă toate valorile observate.

În exemplul luat, folosind un interval de 5 bucăţi, pot apărea mai multe variante de formare a intervalelor de grupare şi de repartizare a unităţilor pe grupe.

Astfel, în varianta de grupare : 120—125; 125—130; 130—135; 135—140; 140—145, limita superioară a fiecărui interval se repetă ca limita inferioară a intervalului următor, deci gruparea este cu variaţie continuă de la o grupă la alta. Acest procedeu de grupare prezintă dificultatea că o variantă a caracteristicii de grupare poate fi inclusă în două grupe alăturate. De pildă, muncitorul care a realizat o producţie de 125 bucăţi poate fi inclus atît în grupa I cît şi în grupa a II-a. Pentru a evita o astfel de situaţie se întocmeşte cîte o notă în care se precizează ce limită trebuie inclusă în interval. Folosind cele două posibilităţi se obţin cele două variante de grupare.

Scanned with OKEN Scanner

Varianta I

grupe după mărimea producţiei obţinute	Nr. muncitorilor
120—125	4
125—130	5
130—135	8
135—140	5
140—145	4

Notă: limita inferioară inclusă în interval

Varianta II

grupe după mărimea producţiei obţinute	Nr. muncitorilor
120—125	7
125—130	3
130—135	10
135—140	2
140—145	4

Notă: limita superioară inclusă în interval

Comparînd frecvenţele pentru cele două variante se observă că ele diferă tocmai datorită faptului că au fost suficiente valori ale caracteristicii egale cu una din limitele intervalelor de grupare.

Pentru a elimina această dificultate se fac grupări cu variaţia discontinuă în care limita inferioară a intervalului următor este deplasată cu o unitate de măsură faţă de limita superioară a intervalului precedent.

Variantele III şi IV nu sînt decît aceleaşi grupări, dar cu limite precizate.

Varianta III

grupe după mărimea producţiei obţinute	Nr. muncitorilor
120—124	4
125—129	5
130—134	8
135—139	5
140—144	4

Varianta IV

grupe după mărimea producţiei obţinute	Nr. muncitorilor
121—125	7
126—130	3
131—135	10
136—140	2
141—145	4

Dacă se compară frecvenţele în cele patru variante de prezentare a rezultatelor grupării se constată că, de fapt, sînt două situaţii: gruparea cu limita inferioară inclusă în interval (variantele I şi III) şi gruparea cu limita superioară inclusă în interval (variantele II şi IV).

Alegerea limitelor care se includ în interval se face în funcţie de datele ce urmează să fie centralizate, astfel încît să nu rămînă nici o unitate în afara grupării şi să evite pe cît posibil asimetria de repartiţie a frecvenţelor [1].

De precizat, că atunci cînd există o singură variantă care apare cu o valoare care are o abatere foarte mare faţă de toate celelalte valori individuale este indicat să se facă o grupă marginală pentru varianta respectivă.

[1] Noţiunea de asimetrie şi măsurarea ei statistică sînt precizate în capitolul „Mărimile medii, indicatorii variaţiei şi asimetriei".

Scanned with OKEN Scanner

De exemplu, dacă se grupează municipiile şi oraşele ţării după numărul populaţiei, atunci este indicat ca amplitudinea variaţiei să se stabilească fără municipiul Bucureşti şi la sfîrşitul grupării să se treacă separat acest municipiu.

La determinarea mărimii intervalului de grupare, în special pentru caracteristicile statistice cu tendinţe de variaţie sistematică şi număr suficient de mare de observaţii, se poate folosi formula lui H.D. Sturges:

$$K = \frac{x_{max} - x_{min}}{1 + 3{,}322 \log N}$$, în care N reprezintă numărul total al observaţiilor

$$K = \frac{21}{1 + 3{,}322 \times \log 26} = 3{,}68 \approx 4.$$

Mărimea intervalului de grupare. În general, mărimea intervalului de grupare se obţine făcînd diferenţa fie între două limite inferioare alăturate fie între două limite superioare alăturate (variantele III şi IV), fie între limita superioară şi limita inferioară a aceluiaşi interval (variantele I şi II). Pentru intervalele cu variaţie discontinuă trebuie să se adauge şi unitatea cu care s-a făcut separarea intervalelor.

În publicaţiile statistice se pot întîlni cazuri de grupări în care limita inferioară a primului interval şi limita inferioară a ultimului interval nu sînt precizate. Se spune despre această grupare că foloseşte intervale deschise. Pentru a putea aplica procedeele de calcul statistic este necesară închiderea intervalelor, plecînd de la ipoteza că ele ar fi egale cu intervalele alăturate.

Deci, primul interval se închide scăzînd din limita superioară mărimea intervalului grupei a doua şi obţinînd astfel limita sa inferioară, iar ultimul interval se închide prin stabilirea limitei sale superioare, adăugînd la limita sa inferioară mărimea intervalului penultim. Dacă mărimea intervalului grupei a doua este mai mare decît limita superioară a primei grupe, atunci se consideră în mod convenţional valoarea minimă a caracteristicii ca fiind egală cu 0 sau 1.

3.4.4. Funcţiile grupării statistice

Principalele funcţii de cunoaştere a grupărilor folosite în studierea fenomenelor şi proceselor social-economice sînt:

— determinarea structurii colectivităţii cercetate pe tipuri calitative diferenţiate în cadrul aceleiaşi colectivităţi;

— sesizarea mutaţiilor produse în structura colectivităţii statistice, pe plan teritorial şi în dinamică, ca urmare a modificării continue a condiţiilor în care acţionează legile dezvoltării sociale;

— surprinderea tendinţelor de manifestare a variaţiei fenomenelor studiate de statistică;

Scanned with OKEN Scanner

— stabilirea și interpretarea legăturilor dintre fenomene și a factorilor care le influențează.

Un exemplu, în care se pot urmări principalele funcții ale grupării este cel prezentat în tabelul 3.1.

În acest tabel caracteristica de grupare o constituie numărul de muncitori ocupați în întreprinderile de subordonare republicană. Gruparea este cu intervale neegale și deschise la ambele capete, confirmînd o amplitudine mare a variației caracteristicii de grupare.

Tabelul 3.1

Structura industriei de stat, de subordonare republicană a R.S. România în anii 1960 și 1975 după numărul muncitorilor

Grupe de întreprinderi după numărul muncitorilor	Nr. întreprinderilor		Nr. muncitorilor		Prod. globală industrială	
	1960	1975	1960	1975	1960	1975
0	1	2	3	4	5	6
pînă la 200	19,4	11,5	2,7	0,9	4,0	1,9
201—500	31,8	16,6	12,6	2,7	17,5	5,9
501—1 000	24,7	22,0	20,6	10,7	20,2	10,0
1 001—2 000	15,3	25,6	25,0	23,6	21,6	19,1
2 001—3 000	5,1	12,3	14,2	19,1	14,4	16,7
3 001—5 000	2,1	8,1	9,8	19,2	7,8	18,7
peste 5001	1,6	3,9	15,6	22,8	14,5	27,7
Total	100,0	100,0	100,0	100,0	100,0	100,0

Sursa: Anuarul Statistic al R.S. România, pe anii 1961, p. 132—133 și 1976, p. 86—87.

În coloanele 1 și 2 sînt centralizate frecvențele de apariție a caracteristicii exprimate sub formă de greutăți specifice față de total.

În următoarele coloane sînt centralizate, condiționat de variația caracteristicii de grupare, mărimea numărului de muncitori și valoarea producției globale industriale.

Analiza succintă a datelor din tabel arată: în 1960 frecvența cea mai mare o aveau întreprinderile în care lucrau între 201—500 muncitori, iar în 1975 frecvența cea mai mare o au întreprinderile în care lucrează între 1 001—2 000 muncitori.

Mutațiile intervenite în structura colectivității în funcție de valorile caracteristicii de grupare se pot urmări comparînd ponderile de la o perioadă la alta. Se poate observa că în 1975 ponderea primelor trei grupe a scăzut, în timp ce la celelalte grupe s-a manifestat o tendință de creștere mai accentuată la grupa centrală. În același sens se realizează mutațiile de structură în distribuția numărului de muncitori și a producției globale obținute.

Aceste deplasări de structură, în sensul creșterii ponderii către valorile superioare ale caracteristicii de grupare, atestă legătura dintre numărul muncitorilor — factor extensiv — și mărimea producției industriale. Se poate constata, pe cei doi ani comparați, la fiecare grupare tendința de

Scanned with OKEN Scanner

apropiere a valorii ponderilor calculate pentru numărul muncitorilor şi pentru producţia globală. Dacă ponderile ar fi fost egale între ele, ar fi însemnat că numărul de muncitori era singurul factor care determină mărimea producţiei. Producţia, depinzînd şi de alţi factori, înseamnă că cu cît aceştia au o influenţă mai mare, cu atît abaterile dintre ponderi sînt mai semnificative.

Pentru a putea caracteriza mai bine tendinţa de centralizare şi concentrare a producţiei în întreprinderile industriale de subordonare republicană, se va folosi o *grupare repetată* [1], în care se va restrînge numărul de grupe. S-ar putea considera întreprinderile împărţite în patru grupe, în funcţie de mărimea numărului de muncitori ocupaţi: întreprinderi mici, întreprinderi mijlocii, întreprinderi mari, întreprinderi foarte mari.

Datele fiind exprimate în % faţă de total — deci faţă de aceeaşi bază — în interiorul aceleiaşi coloane se pot face operaţii de însumare în vederea restrîngerii grupelor.

Tabelul 3.2

Structura industriei de stat, de subordonare republicană a R.S. România, în anii 1960 şi 1975 după numărul muncitorilor

(în procente)

Grupe de întrep. după nr. muncitorilor	Nr. întreprinderilor		Nr. muncitorilor		Prod. globală industrială	
	1960	1975	1960	1975	1960	1975
0	1	2	3	4	5	6
pînă la 500	51,2	28,1	15,3	4,6	21,5	7,8
501—2 000	40,0	59,9	45,6	34,3	41,8	22,1
2 001— 5000	7,2	20,4	23,5	38,3	22,2	35,4
peste 5 000	1,6	3,6	15,6	22,8	14,5	27,7
Total	100,0	100,0	100,0	100,0	100,0	100,0

Regrupînd datele într-un număr mai mic de grupe, gradul de concentrare şi legătura dintre cele două caracteristici apar mai clar. De asemenea, structura pe tipuri calitative şi mutaţiile produse între anii 1960—1975 ies mai clar în evidenţă pe baza datelor noii grupări. Este evident faptul că, deşi întreprinderile mari şi foarte mari au o pondere mică în industria de subordonare republicană, totuşi ele aveau în 1975 o pondere de peste 60% în ceea ce priveşte numărul muncitorilor şi producţia globală industrială. Aceasta confirmă faptul că întreprinderile mari pot asigura şi un grad mai înalt de tehnicitate a producţiei şi, respectiv, un nivel de productivitate mai ridicat şi care se regăsesc într-o valoare mai mare a producţiei.

Din cele prezentate pînă aici apare clar că metoda grupării este metoda de bază prin care se asigură separarea pe clase omogene în studierea statistică a fenomenelor şi proceselor social-economice de masă.

[1] Pe baza datelor din tabel, care sînt rezultatul unei grupări, se obţin noi grupe — deci o regrupare.

Scanned with OKEN Scanner

3.5. FOLOSIREA INDICATORILOR STATISTICI ABSOLUȚI ȘI DERIVAȚI ÎN PREZENTAREA ȘI ANALIZA REZULTATELOR CENTRALIZĂRII ȘI GRUPĂRII

Efectuînd sistematizarea datelor statistice cu ajutorul centralizării și grupării (centralizarea simplă și pe grupe) se obțin indicatorii totalizatori parțiali și generali cunoscuți sub denumirea de indicatori statistici absoluți. Fiind obținuți prin însumarea datelor individuale sau parțiale înregistrate pentru aceeași caracteristică, înseamnă că sînt întotdeauna exprimați în unități concrete de măsură (bucăți, kg, m etc.) sau în cazul cînd sînt caracteristici ale căror valori nu pot fi însumabile direct, se folosesc unități naturale convenționale sau valorice. De aici rezultă că indicatorii absoluți, denumiți și mărimi absolute, se caracterizează prin faptul că atît conținutul cît și forma lor de exprimare sînt concrete. De exemplu, nivelurile producției realizate într-o întreprindere, ramură sau pe economia națională sînt mărimi statistice la fel de concrete ca și mărimea producției realizată de fiecare muncitor în parte. Diferența dintre aceste mărimi statistice constă numai în gradul de cuprindere al unităților colectivității în calculul indicatorilor respectivi.

În cadrul unei cercetări concrete, indicatorii absoluți se pot constitui sub forma unui sistem de indicatori care au proprietatea de a fi variabili în timp, în spațiu sau de la o grupă la alta și sînt considerați ca bază a cunoașterii statistice. Fiind legați de forma lor concretă de exprimare, ei nu pot fi folosiți direct în analiza statistică decît pentru aceeași caracteristică.

Pentru a putea analiza relațiile de interdependență dintre mai mulți indicatori absoluți este necesar să se calculeze indicatorii statistici derivați.

Indicatorii statistici derivați sînt aceia care se obțin din prelucrarea indicatorilor absoluți, prin aplicarea variatelor metode și procedee de calcul statistic și fac posibilă analiza trăsăturilor esențiale ale fenomenelor coexistente în spațiu, sau existente în unități de timp diferite.

Cu ajutorul indicatorilor statistici derivați se pot caracteriza : raporturile cantitative dintre diferitele caracteristici statistice, dintre diferitele părți ale unei colectivități sau dintre fenomenele ce se găsesc într-un anumit grad de interdependență ; valorile tipice ce se formează în mod obiectiv în cadrul aceleiași perioade de timp sau în dinamică ; gradul și forma de variație a caracteristicilor cercetate ; legăturile de interdependență dintre fenomene ; tendința obiectivă de dezvoltare a fenomenelor și dinamica lor ; rolul și contribuția diferiților factori la formarea mărimii unui fenomen complex etc.

Pentru rezolvarea acestor sarcini de cunoaștere se folosesc în statistica social-economică indicatorii derivați calculați sub formă de : mărimi relative, mărimi medii, indicatori de variație și corelație, indici, ecuații de estimare etc.

Metodologia de calcul al acestor indicatori derivați este diferită și va fi prezentată în capitolele ce urmează.

Trăsătura comună a acestor indicatori derivați constă în faptul că ei se formează ca indicatori cu caracter abstract, exprimînd relații la ni-

Scanned with OKEN Scanner

velul unei grupe sau pe total, chiar dacă uneori ei se exprimă în unităţi concrete de măsură. De exemplu, nivelul mediu al productivităţii muncii într-o ramură este un indicator derivat, care se poate măsura în unităţi naturale, natural-convenţionale, valorice sau de timp de muncă. Caracterul abstract al acestui indicator derivă din faptul că, fiind o mărime sintetică, mărimea sa poate să nu fie întîlnită ca valoare concretă la nici un muncitor din colectivitatea respectivă sau, chiar dacă apare, aceasta este cu totul întîmplătoare.

Dacă forma acestor indicatori este abstractă, conţinutul lor trebuie să fie concret, exprimînd tendinţa de manifestare statistică a unor fenomene din realitatea social-economică. Conţinutul concret al acestor indicatori derivaţi este asigurat de gradul de omogenitate a unităţilor la care se aplică metodele de calcul şi analiză statistică.

Cunoaşterea statistică, al cărei obiect îl constituie descoperirea legităţilor de apariţie şi dezvoltare a fenomenelor de masă — fenomene complexe multifactoriale în interiorul cărora legile de probabilitate au un cîmp larg de acţiune — nu se poate realiza dacă nu se bazează pe folosirea combinată a unor indicatori absoluţi şi derivaţi, veridici, calculaţi la diferite nivele de prelucrare statistică a datelor primare.

Scanned with OKEN Scanner

PREZENTAREA DATELOR STATISTICE

Rezultatele prelucrării datelor observării de masă se prezintă sub forme specifice statisticii — tabele, serii, grafice — în care relaţiile dintre fenomenele studiate apar într-o succesiune logică, corespunzătoare relaţiilor obiective existente. Această sistematizare a informaţiilor obţinute face posibilă interpretarea statistică a formelor de manifestare a fenomenelor şi permite alegerea corectă a metodologiei de calcul al indicatorilor statistici. În tabele, serii şi grafice se pot întîlni fie indicatori statistici absoluţi, fie indicatori statistici derivaţi. Cu ajutorul lor se pot caracteriza atît grupele care compun colectivitatea cît şi întregul ansamblu.

4.1. TABELE STATISTICE

Tabelele statistice reprezintă o formă de sistematizare a unui ansamblu de relaţii cantitative despre fenomenele studiate, folosind o reţea de linii paralele, orizontale şi verticale în care se înscriu indicatorii obţinuţi prin prelucrare. Ele se pot folosi în dublu scop :

— pentru prezentarea rezultatelor cercetării în cadrul analizei indicatorilor absoluţi şi derivaţi, care s-au obţinut în diferitele faze ale prelucrării datelor statistice ;

— pentru sistematizarea datelor statistice în vederea aplicării procedeelor de calcul al indicatorilor derivaţi.

Oricare ar fi destinaţia tabelelor statistice, ele trebuie să fie întocmite după anumite principii, respectînd unele reguli de conţinut şi de formă.

4.1.1. Elementele tabelului statistic

Elementele de conţinut ale unui tabel statistic se referă la subiectul şi predicatul tabelului.

Subiectul tabelului îl constituie colectivitatea statistică la care se referă datele prezentate.

Predicatul tabelului este format din sistemul de caracteristici pentru care s-a făcut centralizarea datelor şi care permite caracterizarea statistică a fenomenelor studiate în condiţii specifice de timp şi spaţiu.

Elementele ce ţin de forma unui tabel statistic se regăsesc, în principal, în macheta tabelului. Macheta tabelului statistic este formată din titlul general, titlurile interioare şi reţeaua de rînduri şi coloane.

Reţeaua tabelului este formată din linii paralele, orizontale şi verticale, necesare înscrierii ordonate a datelor statistice. Intervalul între două linii orizontale formează un rînd, iar cele cuprinse între două linii verticale formează o coloană. Întretăierea dintre rînduri şi coloane formează rubricile tabelului în care se înscriu valorile numerice ale indicatorilor statistici.

Scanned with OKEN Scanner

Titlul general se trece în partea de sus a tabelului şi trebuie să fie clar, complet şi concis, atrăgînd atenţia asupra relaţiilor ce trebuie să fie analizate în legătură cu subiectul prezentat. În titlul general trebuie să se precizeze caracteristicile de timp şi spaţiu la care se referă datele prezentate.

Titlurile interioare se înscriu în capetele coloanelor sau ale rîndurilor şi se referă la elementele componente ale subiectului (de regulă, în capetele rîndurilor) şi ale predicatului (de regulă, în capetele coloanelor).

Notele explicative însoţesc un tabel statistic atunci cînd este necesar să se precizeze anumite aspecte legate de sursa de informaţie sau de procedeele de culegere şi prelucrare a datelor prezentate.

4.1.2. Felurile tabelelor statistice

Tabelele statistice sînt extrem de variate, întocmindu-se în funcţie de scopul prelucrării sau al analizei statistice. Cel mai frecvent, tabelele statistice sînt folosite pentru prezentarea rezultatelor centralizării simple sau pe grupe, servind pentru caracterizarea independentă sau interdependentă a fenomenelor. În acest scop, tabelul poate fi prelucrat numai după subiect, uneori după predicat sau combinînd prelucrarea subiectului cu cea a predicatului. Aceasta înseamnă că s-a folosit o grupare combinată după variaţia a cel puţin două caracteristici. Tabelele statistice des întîlnite sînt : tabele simple, tabele pe grupe ; tabele combinate ; tabele cu dublă intrare ; tabele de asociaţie.

Tabelele simple sînt cele în care se prezintă indicatorii statistici ai unităţilor statistice la care se referă datele, ordonaţi după următoarele criterii : cronologic, teritorial şi organizatoric. Întocmirea unor astfel de tabele nu ridică probleme deosebite în legătură cu modul de prezentare a indicatorilor, ordonarea făcîndu-se în funcţie de scop.

Probleme legate de modul de întocmire a unui tabel statistic se pun atunci cînd colectivitatea este împărţită în grupe după variaţia uneia sau a mai multor caracteristici sau cînd şi predicatul se centralizează pe grupe. Atunci este necesar să se întocmească un tabel în care să se poată urmări cu uşurinţă succesiunea logică a informaţiilor prezentate, condiţionată de variaţia caracteristicii de grupare.

Tabelul pe grupe se foloseşte cînd se aplică gruparea simplă şi se centralizează frecvenţele şi valorile caracteristicilor care se găsesc într-o relaţie de dependenţă faţă de variaţia caracteristicii de grupare. În acest caz, valorile caracteristicilor dependente (y, z, v) apar ca sume condiţionate ($\Sigma y/x$; $\Sigma z/x$; $\Sigma v/x$) de variaţia factorului de grupare (x).

Deci se foloseşte pentru grupare caracteristica x_i, cu valorile individuale :

$$x_1 x_2 \ldots x_i \ldots x_m.$$

În acest tabel subiectul este reprezentat prin grupele formate pe baza caracteristicii de grupare x, iar predicatul este format din frecvenţele de apariţie ale diferitelor variante (x_1, $x_2 \ldots x_m$) şi din sumele parţiale ale valorilor înregistrate pentru caracteristicile y, z, v, condiţionate de varia-

Scanned with OKEN Scanner

Modelul tabelului pe grupe

Grupe de unităţi după variaţia caracteristicii x	Numărul unităţilor (f)	Valorile centralizate ale caracteristicilor		
		y	z	v
A	1	2	3	4
x_1	f_1	y/x_1	z/x_1	v/x_1
x_2	f_2	y/x_2	z/x_2	v/x_2
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_i	f_i	y/x_i	z/x_i	v/x_i
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_m	f_m	y/x_m	z/x_m	v/x_m
Total	$\sum_{i=1}^{m} f_i$	$\sum_{i=1}^{m} y/x_i$	$\sum_{i=1}^{m} z/x_i$	$\sum_{i=1}^{m} v/x_i$

ţia valorilor variabilei x. Deci, tabelul poate fi folosit pentru următoarele scopuri statistice:

— caracterizarea independentă a gradului şi formei de variaţie a caracteristicii x;

— interpretarea legăturilor dintre variaţia caracteristicii de grupare şi variaţia caracteristicilor care formează predicatul tabelului;

— pentru aplicarea metodelor de calcul ale corelaţiei statistice.

Tabelul combinat se foloseşte cînd subiectul se prezintă prelucrat după variaţia a cel puţin două caracteristici de grupare (x, y) şi predicatul este format din valorile centralizate ale variabilelor dependente (z, v) de factorii de grupare.

Aici se folosesc pentru grupare:

— variabila x cu valorile: $x_1, x_2, \ldots x_i \ldots x_m$;

— variabila y cu valorile: $y_1, y_2, \ldots y_j \ldots y_p$.

În acest tabel subiectul este format din grupele în care a fost împărţită colectivitatea în funcţie de valorile perechi ale celor două caracteristici x, y, iar predicatul tabelului este format din frecvenţele comune ale valorilor perechi xy şi sumele parţiale ale caracteristicilor z, v condiţionate concomitent de valorile lui x şi y (vezi tabelul 4.2).

Folosind indicatorii din tabel se pot rezolva următoarele aspecte:

— gradul şi forma de variaţie a primei caracteristici de grupare, considerată ca o variabilă independentă ale cărei niveluri individuale sînt prezentate în coloana A, iar frecvenţa în rîndurile totalurilor parţiale calculate pe grupe în coloana 1;

— gradul şi forma de variaţie a valorilor caracteristicii y condiţionată de valorile caracteristicii x;

— legăturile dintre valorile perechi ale caracteristicilor de grupare şi valorile centralizate ale variabilelor z, v;

Scanned with OKEN Scanner

— aplicarea metodelor de calcul al corelației statistice dintre cele patru caracteristici.

Tabelul cu dublă intrare se folosește atunci cînd colectivitatea a fost împărțită în grupe după variația a două caracteristici (x, y) și au fost centralizate numai frecvențele de apariție ale valorilor x, y.

Tabelul 4.2

Modelul tabelului combinat

Caracteristica primară de grupare: x	Caracteristica secundară de grupare: y	Frecvențele corespunzătoare valorilor xy	Valorile centralizate ale caracteristicilor	
			z	v
A	B	1	2	3
x_1	y_1	f_{11}	z/x_1y_1	v/x_1y_1
	y_j	f_{1j}	z/x_1y_j	v/x_1y_j
	y_p	f_{1p}	z/x_1y_p	v/x_1y_p
	Total gr. I	$\sum_{j=1}^{p} f_{1j}$	$\sum_{j=1}^{p} z/x_1y_j$	$\sum_{j=1}^{p} v/x_1y_j$
x_i	y_1	f_{i1}	z/x_iy_1	v/x_iy_1
	y_j	f_{ij}	z/x_iy_j	v/x_iy_j
	y_p	f_{ip}	z/x_iy_p	v/x_iy_p
	Total gr. i	$\sum_{j=1}^{p} f_{ij}$	$\sum_{i=1}^{p} z/x_iy_j$	$\sum_{j=1}^{p} v/x_iy_j$
x_m	y_1	f_{m1}	z/x_my_1	v/x_my_1
	y_j	f_{mj}	z/x_my_j	v/x_my_j
	y_p	f_{mp}	z/x_my_p	v/x_my_p
	Total gr. m	$\sum_{j=1}^{p} f_{mp}$	$\sum_{j=1}^{p} z/x_my_j$	$\sum_{j=1}^{p} v/x_my_j$
Total general		$\sum_{i=1}^{m} f_i$	$\sum_{i=1}^{m} z/x_i$	$\sum_{i=1}^{m} v/x_i$

Scanned with OKEN Scanner

Într-un tabel cu dublă intrare grupele formate după variația caracteristicii x reprezintă elementele componente ale subiectului, iar grupele formate după variația caracteristicii y, elementele componente ale predicatului. În rubricile tabelului se trec frecvențele valorilor x, y.

Într-o formă generalizată tabelul cu dublă intrare pentru combinarea caracteristicilor x, y se prezintă în tabelul 4.3.

Tabelul 4.3

Modelul tabelului cu dublă intrare

x \ y	y_1	$y_2 \dots$	$y_j \dots$	y_p	Frecvențe după variabila x
x_1	f_{11}	$f_{12} \cdots$	$f_{1j} \cdots$	f_{1p}	f_{x_1}
x_2	f_{21}	$f_{22} \cdots$	$f_{2j} \cdots$	f_{2p}	f_{x_2}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
x_i	f_{i1}	f_{i2}	f_{ij}	f_{ip}	f_{x_i}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
x_m	f_{m1}	f_{m2}	f_{mj}	f_{mp}	f_{x_m}
Frecvențe după variabila y	f_{y_1}	$f_{y_2} \cdots$	$f_{y_j} \cdots$	f_{y_p}	$\sum_{i=1}^{m} f_{x_i} = \sum_{j=1}^{p} f_{y_j} = N$

Rezultă că într-un tabel cu dublă intrare, unitățile la care s-a făcut înregistrarea datelor se distribuie atît după variația lui x, cît și după variația lui y, pentru care, deci, numărul total al unităților observate (N) este egal atît cu suma frecvențelor după x, $\sum_{i=1}^{m} f_{x_i}$ cît și cu cele după y, $\sum_{j=1}^{p} f_{y_j}$.

Pe baza tabelului cu dublă intrare se pot caracteriza :

— forma de variație a variabilei x luată independent, folosind distribuția sa marginală (vezi capetele rîndurilor și ultima coloană) ;

— forma de variație a variabilei y luată independent, folosind distribuția sa marginală (vezi capetele coloanelor și ultimul rînd) ;

— distribuțiile condiționate ale caracteristicii x în funcție de y și ale caracteristicii y în funcție de x.

Tabelul de asociație se folosește pentru a putea prezenta legătura dintre două caracteristici alternative. În tabelul de asociație nu sînt decît două variante x_1, x_2 pentru grupele formate pe baza variației subiectului și y_1, y_2 pentru grupele formate pe baza variației predicatului. În rubricile tabelului se trec frecvențele comune ale combinării succesive a valorilor celor două variante (a, b, c, d, iar N reprezintă numărul unităților observate).

Scanned with OKEN Scanner

Acest tabel de asociaţie poate fi folosit şi pentru caracteristici nealternative transformate în variabile alternative. De exemplu, muncitorii care au producţie individuală sub sau peste nivelul mediu calculat pe întreaga

Tabelul 4.4

Modelul tabelului de asociaţie

x \ y	y_1	y_2	Total
x_1	a	b	a + b
x_2	c	d	c + d
Total	a + c	b + d	(a + c) + (b + d) = = (a + b) + (c + d) = N

întreprindere şi muncitori care au o retribuţie sub şi peste nivelul mediu al retribuţiei.

4.1.3. Reguli pentru întocmirea şi interpretarea tabelelor statistice

Tabelele statistice trebuie să fie astfel întocmite încît să constituie într-adevăr mijloace de sistematizare a unor date, de obicei extrem de variate între ele. Pentru aceasta este necesar să se respecte anumite reguli privitor la modul concret de întocmire a unui tabel statistic.

Stabilirea subiectului şi predicatului tabelului se face în funcţie de scopul sistematizării datelor statistice.

Pornind de la cunoaşterea scopului pentru care se elaborează un tabel statistic, se stabilesc subiectul şi predicatul tabelului, avînd grijă ca ele să se regăsească distinct în macheta tabelului. Subiectul trebuie astfel prezentat încît să rezulte clar care sînt grupele sau subgrupele colectivităţii (delimitată în timp şi spaţiu) la care se referă valorile indicatorilor care au fost calculaţi. Predicatul tabelului trebuie ales astfel încît să corespundă direct necesităţii de analiză statistică, cuprinzînd caracteristicile dependente de variaţia caracteristicilor de grupare a unităţilor colectivităţii care constituie subiectul tabelului. Predicatul tabelului trebuie să fie trecut astfel încît să se poată urmări cu uşurinţă ordinea legăturilor dintre caracteristicile care-l compun. Această cerinţă este satisfăcută dacă caracteristicile studiate sînt prezentate sub formă de sistem, în cadrul căruia se pot interpreta relaţiile pe grupe, între grupe şi faţă de total.

Indicatorii statistici avînd, de regulă, caracter concret, ei se exprimă în diferite unităţi de măsură. Dacă unitatea de măsură este comună pentru toţi indicatorii din tabel, atunci ea se înscrie în afara reţelei tabelului, de obicei, în dreapta sus sau în titlul general al tabelului. Dacă unităţile de măsură sînt diferite, atunci pentru fiecare indicator, sub titlul interior (al coloanei sau rîndului) se înscrie unitatea de măsură respectivă.

Scanned with OKEN Scanner

Rețeaua unui tabel devine un tabel statistic atunci cînd toate rubricile lui sînt completate cu indicatori statistici. În cazul în care pentru o grupă, într-o unitate de timp și spațiu, nu a existat fenomenul respectiv, se trage o linie (—), iar dacă se știe că fenomenul a existat, dar lipsesc datele, atunci se trage o linie întreruptă formată din trei puncte (...).

Tabelele statistice trebuie să fie ușor de interpretat. Pentru aceasta este necesar să cuprindă numai informații strict necesare caracterizării statistice a fenomenelor studiate. În cazul folosirii unui număr mai mare de grupe sau caracteristici este indicat să se separe informațiile în mai multe tabele parțiale și indicatorii sintetici să se reprezinte într-un tabel centralizator. Dacă sînt necesare tabele mai ample sau tabele pentru calcularea unor indicatori statistici, acestea trebuie să fie date ca tabele anexe. Dacă valoarea indicatorilor este mare — cu cifre greu de reținut — este indicat să se exprime indicatorii în mii sau milioane de unități. În acest caz trebuie să se indice în tabel modul de prezentare a valorilor indicatorilor statistici, iar datele se vor rotunji astfel : pentru exprimarea în mii unități se va renunța la tot ce este sub 500 și se va rotunji în plus cu o unitate pentru tot ce depășește 500. De exemplu 565 865 se va trece 566 mii, iar pentru 565 465 se va trece 565 mii.

Orice tabel statistic trebuie să prezinte corect sursa de informație, arătînd locul și timpul la care se referă datele, precum și cine furnizează informația.

Notele explicative sînt necesare numai atunci cînd trebuie să fie atrasă atenția asupra modului de culegerea sau de prelucrare a datelor din același tabel.

4.2. SERIILE STATISTICE

Seriile statistice se obțin în urma aplicării metodei grupării folosind una sau mai multe caracteristici de grupare. De cele mai multe ori seriile statistice trebuie să fie extrase din tabelele de prezentare a rezultatelor prelucrării datelor statistice.

Seria statistică este o corespondență între două șiruri de date statistice în care primul șir reprezintă variația caracteristicii de grupare, iar cel de-al doilea șir rezultatul centralizării frecvențelor de apariție sau a valorilor unei alte caracteristici cu care se corelează.

Seria statistică poate fi considerată, astfel, ca o funcție matematică în care valorile centralizate ale frecvențelor sau ale caracteristicilor sînt valori dependente (y) în funcție de valorile caracteristicii de grupare (x).

4.2.1. Feluri de serii statistice

După posibilitatea de caracterizare a fenomenelor, seriile statistice pot fi : serii statistice independente și serii statistice condiționate.

Seriile statistice independente sînt acelea care rezultă dintr-o grupare simplă, iar *seriile statistice condiționate* sînt acelea obținute dintr-o grupare combinată.

Scanned with OKEN Scanner

Seriile statistice independente se mai numesc şi serii unidimensionale, iar cele condiţionate — multidimensionale.

După conţinutul caracteristicii de grupare, seriile statistice pot fi: serii statistice de timp (dinamice sau cronologice); serii statistice de spaţiu (teritoriale); serii statistice de distribuţie (de repartiţie).

Seriile statistice de timp (dinamice sau cronologice) sînt acelea în care se prezintă variaţia unei caracteristici în funcţie de timp.

Relaţia matematică în acest caz este: $y_i = f(t_i)$, în care:

y_i reprezintă valorile caracteristicii a cărei variaţie se studiază;

t_i — variabila de timp.

După timpul la care se referă datele prezentate pot fi: serii dinamice de intervale şi serii dinamice de momente.

Seriile dinamice de intervale sînt acelea în care valorile caracteristicii studiate sînt rezultatele diferitelor activităţi depuse în cadrul unui proces social-economic pe luni, trimestre etc., ca de exemplu producţia în unităţi naturale sau valorice, cheltuielile de producţie, numărul născuţilor etc. Cumulînd datele din fiecare unitate de timp se poate obţine valoarea centralizată a caracteristicii respective pe întreaga perioadă. Un exemplu de serie dinamică pe intervale se arată în tabelul 4.5.

Tabelul 4.5

Producţia de fontă a R.S. România între anii 1970—1977

	1970	1971	1972	1973	1974	1975	1976	1977
mii tone	4 210	4 382	4 890	5 713	6 081	6 602	7 415	7 784

Sursa: Anuarul statistic al R.S. România, Bucureşti, D.C.S. 1978, p. 168—169.

Din interpretarea acestei serii de date privind producţia de fontă a României se poate observa tendinţa de creştere continuă, fără însă a putea preciza care este măsura acestei creşteri şi nici forma specifică de dezvoltare pe această perioadă de timp.

Prin seria statistică s-a obţinut numai o sistematizare a datelor pentru etapa care interesează şi trebuie în continuare supusă prelucrării statistice [1].

Seriile statistice de momente se obţin pentru variabile statistice pentru care este raţional ca însumarea să se facă la anumite momente, ca, de exemplu: numărul de muncitori, valoarea fondurilor fixe, numărul de tractoare, valoarea mijloacelor circulante, stocurile de mărfuri etc.

Dacă, de exemplu, se prezintă datele privind parcul de tractoare de care a dispus agricultura în perioada 1950—1975, se poate constitui fie o serie de momente pe fiecare an, fie o serie de momente luate din cinci ani.

Din datele prezentate în această serie se poate observa că de la un cincinal la altul parcul de tractoare din agricultură a crescut sistematic de la 13 713 (respectiv 16 746 în tractoare convenţionale) în 1950, la 119 533 (respectiv 213 652 în tractoare convenţionale) în 1975. Această creştere este rezultatul politicii de industrializare a ţării care atrage după sine creşterea gradului de mecanizare a proceselor agricole.

[1] Vezi capitolul „Analiza seriilor cronologice".

Scanned with OKEN Scanner

Tabelul 4.6

Parcul de tractoare din agricultură, în perioada 1950—1975 (la sfîrşitul anilor)

Tractoare	1950	1955	1960	1965	1970	1975
A	1	2	3	4	5	6
— în unităţi fizice	13 713	23 033	44 194	81 356	107 290	119 533
— în unităţi convenţionale de 15 C.P.	16 746	30 488	65 290	132 982	184 772	213 652

Sursa: Anuarul Statistic al R.S. România, Bucureşti, D.C.S., 1976, p. 160—161.

Caracteristic pentru seria dinamică de momente este faptul că termenii ei nu se pot cumula deoarece pot exista înregistrări repetate şi valoarea obţinută n-ar avea sens economic. Caracterizarea statistică pentru întreaga perioadă nu se poate face în acest caz printr-un indicator totalizator, ci folosind indicatori derivaţi sub formă de mărimi relative sau medii.

Este necesar ca pentru fiecare serie dinamică să rezulte clar (din titlul ei) dacă datele prezentate se referă la momente sau la perioade de timp, pentru a putea stabili relaţii corecte între ele.

Seriile statistice de spaţiu sînt cele în care centralizarea frecvenţelor sau a valorilor individuale ale caracteristicii studiate se face în funcţie de variaţia teritorial-administrativă.

De regulă, seriile statistice teritoriale sînt folosite pentru sistematizarea informaţiilor statistice pe judeţele ţării. Pe baza seriilor teritoriale se poate face o „ierarhizare" a judeţelor în vederea distribuirii optime a forţelor de producţie pe plan teritorial.

Pentru a studia, de pildă, distribuţia teritorială a metalurgiei feroase (inclusiv extracţia minereurilor feroase) se întocmeşte o serie statistică în care primul şir este format din judeţele ţării în care apare această activitate, iar în al doilea şir se prezintă volumul sau ponderea ramurii respective a judeţului, faţă de totalul ţării.

Deci se pot întocmi serii teritoriale în care indicatorii pot fi absoluţi, relativi şi medii.

Tabelul 4.7

Structura producţiei globale a metalurgiei feroase (inclusiv extracţia minereurilor feroase) pe judeţele producătoare, în anul 1975

Nr. crt.	Judeţul	în % faţă de total	Nr. crt.	Judeţul	în % faţă de total
A	B	1	A	B	1
1.	Brăila	2,2	8.	Hunedoara	24,8
2.	Buzău	2,7	9.	Iaşi	4,8
3.	Caraş-Severin	10,8	10.	Neamţ	5,0
4.	Cluj	8,1	11.	Prahova	0,3
5.	Dîmboviţa	1,8	12.	Suceava	0,2
6.	Galaţi	32,4	13.	Timiş	0,7
7.	Harghita	0,9	14.	Municipiul Bucureşti	5,3
				Total R. S. România	100,0

Sursa: Anuarul Statistic al R.S. România, Bucureşti, D.C.S., 1976, p. 102—105.

Scanned with OKEN Scanner

Pe baza acestei serii de date se poate observa că în anul 1975 numai 13 județe ale țării și Municipiul București erau producătoare în ramura metalurgiei feroase. Dintre acestea ponderea cea mai mare o avea județul Galați (32,4 %) urmat de județul Hunedoara (24,8 %), iar celelalte județe — inclusiv Municipiul București — cu o pondere între 0,2 % (Suceava) și 10,8 % (Caraș-Severin). Pentru a efectua însă o analiză profundă a repartiției teritoriale a producției, a mutațiilor teritoriale intervenite în decursul unei perioade de timp, această serie statistică nu este suficientă. În completare este necesar să se folosească și datele absolute din care s-au calculat acești indicatori statistici, pe mai mulți ani, și să se coreleze cu alte variabile statistice, care influențează dinamica și structura producției de metalurgie feroasă.

Pornind de la funcțiile de cunoaștere care trebuie realizate folosind datele statistice, seriile teritoriale, de regulă, se întocmesc pentru mai multe șiruri de date, privitor la variația caracteristicilor statistice ce se găsesc într-un raport obiectiv de interdependență. Aceste șiruri apar ca serii statistice interdependente, condiționate de teritoriul la care se referă datele.

Seriile statistice de distribuție sînt acelea în care se folosește pentru gruparea datelor o caracteristică atributivă (calitativă sau numerică) iar centralizarea se face pentru frecvențele la care se înregistrează aceeași variantă, sau pentru o altă caracteristică statistică.

Seriile de distribuție după variația unei caracteristici calitative corespund, de regulă, clasificărilor întîlnite în mod curent în statistica de stat. Aceste serii se întocmesc prin centralizarea datelor observării pe grupe înscrise în nomenclatorul elaborat periodic în funcție de modificările intervenite în forma de variație a fenomenelor respective. Printre aceste serii se pot enumera : distribuții pe ramuri ale economiei naționale ; distribuții pe forme de proprietate ; distribuții pe forme de subordonare administrativă ; distribuții pe profesii etc.

De exemplu, dacă se studiază distribuția numărului mediu al personalului, pe ramurile economiei naționale, se întocmește o serie statistică în care caracteristica de grupare o constituie ramura, iar caracteristica studiată—numărul mediu al personalului.

Tabelul 4.8

Numărul mediu al personalului muncitor pe ramuri ale economiei naționale în anul 1975

Ramura	Mii persoane
Industrie	2 802,1
Construcții	736,4
Agricultură	484,2
Silvicultură	51,2
Transporturi	402,2
Telecomunicații	67,9
Circulația mărfurilor	553,2
Gosp. comunală, de locuințe și alte prestări de servicii	286,1
Învățămînt, cultură, artă	387,5
Știință și servire științifică	84,3
Ocrot. sănăt., asist. socială, cultură fizică	262,0
Administrație	64,1
Celelalte ramuri	119,6

Sursa : Anuarul Statistic al R.S. România, 1976, D.C.S., p. 68—69

Scanned with OKEN Scanner

Seria de distribuţie formată după variaţia unei caracteristici numerice se mai numeşte şi serie de variaţie iar, de regulă, al doilea şir este format din frecvenţele corespunzătoare grupelor. În acest caz, pot apărea serii de distribuţii de frecvenţe direct pe variante sau pe intervale de valori.

Următoarea serie este o distribuţie statistică pe variante.

Tabelul 4.9

Distribuţia născuţilor vii, după rangul născutului viu în anul 1975

Rangul	Numărul născuţilor vii
primul	160 144
al doilea	119 186
al treilea	64 844
al patrulea	39 611
al cincelea	14 988
al şaselea	7 558
al şaptelea	4 601
al optulea şi peste	7 199
nedeclaraţi	54

Sursa: Anuarul Statistic al R.S. România, Bucureşti, D.C.S., 1976, p. 32.

Folosind datele din această serie se poate constata că există o tendinţă de concentrare a frecvenţelor către variantele mai mici ale caracteristicii, astfel încît pe măsură ce cresc valorile caracteristicii scad frecvenţele lor de apariţie. Această formă este cunoscută în statistică sub denumirea de distribuţie în formă de „j".

În cazul unei amplitudini mari a variaţiei se folosesc serii statistice pe intervale de grupare (egale sau neegale).

Un exemplu de serie de distribuţie de frecvenţă în care intervalele de grupare sînt egale se prezintă în tabelul 4.10.

Tabelul 4.10

Decedaţii sub 40 ani după grupa de vîrstă, în anul 1975

Grupe de vîrstă	Nr. decedaţilor
A	1
0— 4 ani	17 790
5— 9 ani	1 331
10—14 ani	742
15—19 ani	1 503
20—24 ani	1 877
25—29 ani	1 807
30—34 ani	1 921
34—39 ani	3 300

Sursa: Anuarul Statistic al R.S. România, D.C.S., 1976, p. 34.

În ceea ce priveşte tendinţa de distribuţie a frecvenţelor, se poate observa că ele se concentrează, cu anumite neregularităţi, către valorile extreme ale variabilei de grupare. Seria în care frecvenţele se concentrează

Scanned with OKEN Scanner

către valorile extreme ale caracteristicii a cărei variație se studiază și descresc asimetric pentru valorile centrale este cunoscută în statistică sub denumirea de distribuție în formă de „U".

În seriile precedente, frecvențele au fost exprimate în mărimi absolute rezultate din centralizarea unităților pe grupe. În practica statistică sau în publicațiile de specialitate se întîlnesc deseori serii statistice cu frecvențe relative. Un exemplu, în acest sens, se prezintă în tabelul 4.11.

Tabelul 4.11

Structura industriei după numărul muncitorilor în anul 1976

Numărul muncitorilor	Nr. întreprinderilor (în % față de total)
A	1
pînă la 200 muncitori	10,3
201— 500 muncitori	19,4
501—1 000 „	23,9
1 001—2 000 „	25,0
2 001—3 000 „	10,8
3 001—5 000 „	7,0
peste 5 000 „	3,6

Sursa: Anuarul Statistic al R.S. România, 1977, D.C.S., p. 122—123.

Această serie este o distribuție de frecvențe relative prin interpretarea căreia se poate stabili care era în 1976 structura industriei în funcție de numărul muncitorilor, adică ponderea pe care o avea fiecare grupă de întreprinderi în întreaga colectivitate. Avînd în vedere variația caracteristicii, avem o serie de frecvențe relative cu intervale neegale de grupare. În ceea ce privește forma de repartiție a frecvențelor, se poate observa că aici tendința este de a se concentra către valorile centrale ale caracteristicii de grupare, descrescînd oarecum simetric față de valorile extreme. Această tendință de repartizare a frecvențelor valorilor unei caracteristici statistice este cunoscută sub denumirea de distribuție simetrică sau normală.

Seriile condiționate (repartițiile multidimensionale) se obțin folosind o grupare combinată după două sau mai multe caracteristici statistice care se găsesc în mod obiectiv într-o strînsă interdependență.

Un model de formare a distribuțiilor combinate a două variabile statistice interdependente este acela din tabelul 4.12.

Din acest tabel se pot extrage următoarele serii de distribuție de frecvențe:

— cîte o serie de distribuție pentru fiecare din cele două caracteristici luate independent una de cealaltă; ele formează distribuțiile marginale ale valorilor înscrise în capetele rîndurilor sau coloanelor. În exemplul luat se poate observa că ambele caracteristici prezintă aproximativ aceeași tendință de asimetrie și anume: valorile mai mici au frecvențe mai mari de apariție în comparație cu cele mari;

— 7 serii de distribuție a întreprinderilor după numărul de muncitori, dar condiționate de valoarea fondurilor fixe;

Scanned with OKEN Scanner

— 7 serii de distribuţie a întreprinderilor după valoarea iniţială a fondurilor fixe, dar condiţionate de numărul de muncitori din întreprinderea respectivă.

Tabelul 4.12

Gruparea întreprinderilor de subordonare republicană după mărimea fondurilor fixe şi numărul muncitorilor în anul 1977 (la sfîrşitul anului).

Grupe de întreprinderi după numărul muncitorilor	Grupe de întreprinderi după valoarea iniţială a fondurilor fixe (mii lei)							Număr de întreprinderi în funcţie de numărul muncitorilor
	sub 10 000	10 001 — 40 000	40 001 — 100 000	100 001 — 200 000	200 001 — 500 000	500 001 — 1 000 000	peste 1 000 000	
A	1	2	3	4	5	6	7	8
Pînă la 200	1	67	44	4	—	—	1	122
201 — 500	5	26	81	49	4	4	6	172
501 — 1 000	6	12	98	64	39	39	9	229
1 001 — 2 000	2	9	70	95	91	91	22	307
2 001 — 3 000	—	—	26	47	73	73	20	187
3 001 — 5 000	—	—	6	12	56	56	27	134
peste 5 000	—	—	—	3	9	9	40	86
Nr. de întreprinderi în funcţie de valoarea fondurilor fixe	13	114	325	274	272	114	125	1 237

Sursa: Prelucrat după Anuarul Statistic al R.S. România, D.C.S., 1978, p. 128—129.

Dacă se analizează formele distribuţiilor condiţionate, se poate observa că între cele două variabile există o anumită legătură, adică pe măsură ce creşte valoarea fondurilor fixe într-o întreprindere industrială, creşte şi numărul muncitorilor.

4.2.2. Elemente de analiză a seriilor statistice

La analiza seriilor statistice se folosesc atît repartiţiile empirice, cît şi repartiţiile teoretice.

În general, printr-o repartiţie empirică se înţelege seria care s-a obţinut din centralizarea datelor unei observări statistice, folosind fie frecvenţele absolute, fie frecvenţele relative.

Repartiţiile teoretice sînt cele elaborate pe baza unei ipoteze de repartiţie a frecvenţelor, astfel încît să se poată stabili relaţii matematice bine determinate între valorile variabilei studiate şi frecvenţele lor de apariţie, interpretate ca o funcţie de probabilităţi. Pentru aceste repartiţii s-au stabilit o serie de parametri care le definesc şi care pot fi folosiţi în compararea repartiţiilor empirice cu cele teoretice. Printre repartiţiile teoretice frecvent utilizate în statistica matematică sînt : repartiţia binomială, repartiţia normală şi repartiţia Poisson. Ele sînt studiate în cursul de Statistică matematică şi proprietăţile lor sînt folosite în statistica social-economică la interpretarea fenomenelor luate independent şi interdependent. De asemenea, ele sînt folosite la interpretarea indicatorilor care

Scanned with OKEN Scanner

caracterizează colectivitatea generală atunci cînd nu se dispune decît de date din observări parţiale.

La analiza seriilor empirice trebuie avut în vedere o serie de proprietăţi care se pot întîlni în toate cazurile sau sînt specifice anumitor serii.

Principalele proprietăţi (trăsături caracteristice) ale seriilor statistice sînt : variabilitatea, forma de repartiţie, omogenitatea şi independenţa (interdependenţa) termenilor.

Variabilitatea termenilor unei serii statistice este determinată de faptul că fenomenele de masă pe care le prezintă nu sînt fenomene univoc determinate, ci ele apar ca rezultat al acţiunii combinate a mai multor cauze, unele cu caracter esenţial, altele cu caracter întîmplător, care se manifestă, de regulă, în condiţii individuale diferite.

Modul de acţiune a acestor cauze este diferit : în timp ce acţiunea cauzelor esenţiale are caracter permanent, este prezentă în toate cazurile individuale şi imprimă fenomenului o dezvoltare sistematică, acţiunea cauzelor neesenţiale poate să fie diferită de la o unitate la alta şi se regăseşte în forme variabile de dezvoltare. Gradul şi forma de variaţie a acestor fenomene de masă sînt determinate, deci, de modul de asociere a acţiunii cauzelor esenţiale cu cele întîmplătoare, asociere care are loc în mod diferit în funcţie de condiţiile specifice fiecărui caz individual în parte. Aceasta face ca în modul de formare obiectivă a nivelurilor individuale ale unei caracteristici statistice să *existe o oarecare tendinţă de concentrare către unele valori tipice, care imprimă seriei o anumită formă de distribuţie.*

În consecinţă, statistica are sarcina de a determina valorile tipice care caracterizează seria respectivă, de a explica diferenţele de nivel dintre termenii individuali şi aceste valori tipice, în scopul interpretării formei şi gradului de variaţie ale caracteristicii studiate şi a comparării lor în timp şi spaţiu.

Practica statistică a dovedit că în cazul în care se prezintă date cu privire la aceleaşi caracteristici înregistrate pentru colectivităţi asemănătoare, distribuite în unităţi de timp diferite sau coexistente în spaţiu, seriile de date obţinute prin centralizarea datelor individuale se apropie ca formă de distribuţie. De exemplu, dacă se compară două sau mai multe serii privind distribuţia retribuţiilor muncitorilor dintr-o ramură în diferite ţări cu o structură social-economică asemănătoare, se poate constata o anumită asemănare în forma de repartizare a frecvenţelor, aceste serii diferenţiindu-se numai din punctul de vedere al nivelului de dezvoltare şi al amplitudinii variaţiei.

Această asemănare ca formă de repartiţie se poate explica numai dacă fenomenele sociale se consideră ca fenomene de masă, legic determinate şi dependente de condiţiile materiale în care ele se produc. Dependenţa lor de condiţiile materiale specifice fiecărei unităţi de timp şi spaţiu face însă ca valorile înregistrate pentru aceeaşi caracteristică să fie diferite şi independente una de alta.

Independenţa termenilor se întîlneşte doar în cazul seriilor de repartiţie de frecvenţă sau teritoriale, unde variabila se înregistrează pentru unităţi de observare distincte. În cazul seriilor cronologice, unde datele se referă la aceeaşi unitate de observare statistică şi în succesiunea apariţiei valorilor în timp, termenii seriei sînt interdependenţi între ei.

Scanned with OKEN Scanner

Deşi independenţi (interdependenţi) şi variabili, termenii unei serii statistice trebuie să fie omogeni între ei. *Omogenitatea valorilor* într-o serie statistică apare ca urmare a faptului că în aceeaşi serie nu se pot prezenta decît date cu privire la aceleaşi fenomene prezentate în dinamică, fie în distribuţia sa economico-organizatorică sau teritorial-administrativă.

De aceea este necesar ca la întocmirea unei serii statistice să se verifice în ce măsură unităţile purtătoare ale caracteristicii cercetate sînt omogene între ele, pot fi supuse în continuare prelucrării statistice, în scopul obţinerii unor valori sintetice reprezentative şi care pot fi folosite în comparaţia statistică de la o unitate de timp la alta sau de la o unitate economică la alta.

Dacă în urma analizei statistice se constată că seria nu prezintă omogenitate, înseamnă că în cazul dat colectivitatea este formată din mai multe tipuri calitative şi seria respectivă trebuie desfăşurată în serii componente. În acest ultim caz este necesar să se verifice în ce măsură tendinţa generală de variaţie din întreaga serie se regăseşte în forma de variaţie specifică fiecărei grupe şi să se interpreteze acţiunea sistematică a factorilor esenţiali la nivelul fiecărei grupe şi pe ansamblu. Aceste probleme se rezolvă în statistică cu ajutorul calculului indicatorilor medii, de variaţie şi asimetrie şi a metodologiei de analiză dispersională, care vor fi tratate în capitolele următoare.

4.3. PREZENTAREA GRAFICĂ A DATELOR STATISTICE

Prezentarea grafică presupune folosirea unor suprafeţe, linii sau puncte pentru a reda imaginea unor fenomene care au fost studiate statistic.

Reprezentările grafice sînt folosite în statistica social-economică pentru următoarele scopuri: interpretarea vizuală a raportului de mărime dintre doi sau mai mulţi indicatori statistici; interpretarea structurii şi mutaţiilor de structură, privite în dinamică sau pe plan teritorial; interpretarea densităţii de repartiţie a frecvenţelor fenomenelor înregistrate într-o cercetare statistică concretă; interpretarea formelor de realizare a interdependenţelor dintre două sau mai multe variabile statistice; interpretarea tendinţelor de dezvoltare a fenomenelor studiate în dinamică pentru etapa dată; popularizarea datelor statistice cu privire la gradul de dezvoltare a fenomenelor considerate în dinamică sau pe plan teritorial.

Reprezentările grafice sînt folosite în statistica social-economică fie ca o metodă de prezentare a rezultatelor cercetărilor statistice, fie ca mijloc de alegere a metodelor şi procedeelor de calcul statistic, diferenţiate în raport cu particularităţile de manifestare a fenomenelor studiate.

Reprezentările grafice se aleg, deci, în funcţie de specificul fenomenelor studiate şi al relaţiilor care trebuie să fie evidenţiate prin folosirea lor. Aceasta înseamnă că, la construirea unei reprezentări grafice, trebuie să se folosească în mod adecvat figurile de suprafaţă sau volum, hărţile şi sistemul de axe rectangulare care pot sugera cel mai bine printr-o singură privire proporţiile reale dintre indicatorii prezentaţi.

Scanned with OKEN Scanner

4.3.1. Elementele de bază ale unui grafic

Întocmirea corectă a unui grafic statistic presupune cunoaşterea şi folosirea corespunzătoare a următoarelor elemente: titlul graficului; reţeaua graficului; scara de reprezentare; notele explicative, inclusiv legenda; sursa informaţiilor utilizate.

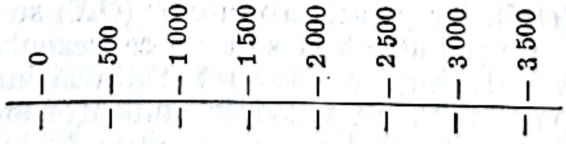
Titlul graficului trebuie să sugereze ce relaţii trebuie să fie interpretate vizual pe baza graficului respectiv. El poate coincide cu titlul tabelului din care s-au extras datele sau se reformulează în raport cu sfera de cuprindere a datelor şi cu scopul pentru care se face reprezentarea grafică. Titlul graficului trebuie să fie scurt, clar şi precis, sugerînd aspectele care sînt evidenţiate în grafic. Titlul se trece, de regulă, deasupra figurii graficului. Dacă graficul face parte dintr-un text, atunci poate fi inclus în fraza care-l precede.

Reţeaua graficului este formată din linii paralele orizontale, verticale sau concentrice care servesc pentru plasarea corectă a punctelor pe grafic. Pentru a se putea scoate în evidenţă liniile sau curbele care reprezintă legătura dintre indicatorii prezentaţi, reţeaua graficului trebuie să fie construită cu linii mai subţiri.

Scara de reprezentare se alege ţinînd seama de ordinul de mărime al indicatorilor de reprezentat, de gradul şi forma de variaţie dintre ei şi de scopul urmărit. Pentru a reda cît mai fidel imaginea reală a fenomenelor studiate, se pot folosi scări uniforme, la care diviziunile de pe suportul scării sînt echidistante între ele, şi scări neuniforme ca, de exemplu, scara logaritmică la care diviziunile suportului sînt fixate în funcţie de valoarea logaritmică a indicatorilor.

Alegerea unităţii de lungime a scării se face în funcţie de spaţiul destinat figurii graficului şi în aşa fel, încît să se surprindă forma reală de variaţie a indicatorilor de reprezentat. Dacă se prezintă corelat mai multe caracteristici statistice, atunci scările de reprezentare trebuie să fie astfel alese, încît să poată cuprinde toate valorile indicatorilor şi să redea într-o formă armonioasă proporţia reală dintre ele.

Pe suportul scării se trec numai valorile care marchează distanţele proporţionale cu unitatea de măsură a scării de reprezentare. De exemplu, dacă pe o axă se consideră că valorile de reprezentat trebuie să respecte proporţia 1 cm = 500 tone şi valoarea minimă este egală cu 530 tone, iar valoarea maximă cu 3 400 tone, atunci pe suportul scării vor apărea valorile:



Valorile reale vor fi plasate pe grafic între diviziunile scării, cu ajutorul reţelei construite în acest scop.

Scările neuniforme se folosesc numai în cazuri speciale, pentru completarea analizei statistice, cînd din graficul construit pe scări uniforme nu reiese destul de clar care este forma de variaţie sau de legătură dintre indicatorii prezentaţi.

Scanned with OKEN Scanner

Notele explicative și legendele se folosesc pentru a putea interpreta corect graficul. Notele explicative apar atunci cînd este necesar să se atragă atenţia asupra aspectelor metodologice ale calculării indicatorilor reprezentaţi, sau asupra modului de prezentare a lor în grafic.

Legendele reprezintă explicarea concisă a semnelor convenţionale, haşurilor şi culorilor folosite.

Sursa de informaţie a datelor din grafic este obligatorie în toate cazurile cînd se folosesc date reale. Ea este trecută în afara figurii graficului şi trebuie să cuprindă toate elementele de identificare a izvorului informaţiei respective. Cînd graficele sînt încadrate într-un raport de analiză, sau într-o lucrare de cercetare ştiinţifică, ca şi titlurile, sursele de informaţie pot să nu apară în mod expres dacă din textul respectiv rezultă aceste elemente.

4.3.2. Principalele tipuri de grafice utilizate în statistica social-economică

În practică, pentru aceleași date statistice se pot folosi uneori mai multe feluri de reprezentări grafice. Alegerea celui mai potrivit tip de grafic se face astfel încît să sugereze, printr-o singură privire, relaţiile obiective dintre indicatorii prezentaţi.

Cele mai frecvente tipuri de grafice sînt : grafice prin coloane şi benzi ; grafice prin figuri geometrice de suprafaţă sau volum ; cronograme (historiograme) ; diagrame radiale (polare) ; diagrame de distribuţie (histogramă, poligon de frecvenţă, curba cumulativă a frecvenţelor) ; cartograme şi cartodiograme ; grafice prin figuri naturale şi simbolice.

Aceste grafice se mai pot grupa şi în funcţie de felul datelor utilizate, sau domeniul de folosire : diagrame ale unor date parţiale sau independente între ele ; diagrame de structură ; graficele seriilor cronologice ; graficele seriilor de distribuţie ; graficele seriilor teritoriale ; graficele de analiza corelaţiei.

În acest capitol, graficele sînt prezentate în funcţie de modul de construire şi nu de domeniul de aplicare.

Graficele prin coloane sînt cele mai frecvent întîlnite în prezentarea datelor statistice. Ele se folosesc în special pentru popularizarea datelor statistice sau a indicatorilor de plan.

Reprezentarea grafică prin coloane presupune folosirea cadranului I din sistemul axelor rectangulare, unde scara de reprezentare se fixează pe axa ordonatelor (OY), iar pe axa absciselor (OX) se construiesc atîtea coloane cu baze egale cîţi indicatori sînt de reprezentat. Coloanele sînt separate între ele la o distanţă egală cu 1/2 din lăţimea lor. Înălţimea coloanelor este proporţională cu valoarea indicatorilor de reprezentat. Punctele cotate pe axa ordonatelor sînt stabilite în funcţie de valoarea maximă a indicatorilor de reprezentat şi a unităţii de lungime a scării.

De exemplu, la o întreprindere care are 3 secţii s-au înregistrat următorii indicatori de îndeplinire a planului ; secţia I—115% ; secţia II—108% ; secţia III—120%.

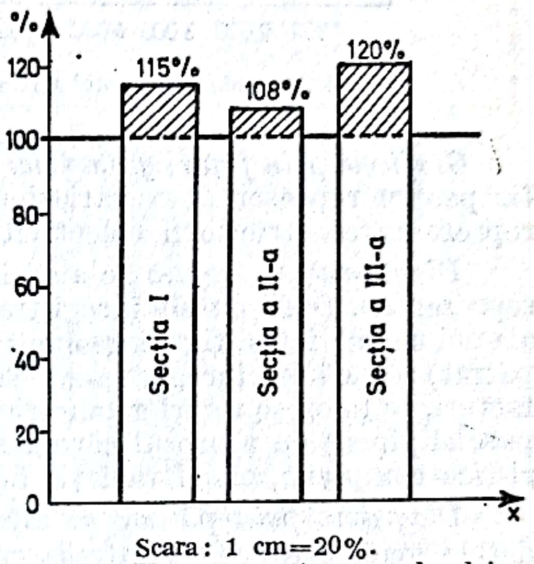
Valoarea maximă de reprezentat este de 120% şi se poate lua ca unitate de lungime a scării : 1 cm = 20%.

Scanned with OKEN Scanner

Pentru a putea compara vizual depăşirea planului, în dreptul realizării integrale (100%) se trage cîte o linie întreruptă în fiecare coloană, iar mărimea procentuală a depăşirii se haşurează şi se trece în legendă (vezi fig. 4.1).

Graficele prin coloane se folosesc, de asemenea, şi în cazul cînd sînt prezentate comparativ datele de la mai multe unităţi teritoriale sau administrative. Ele se folosesc şi în cazul în care se prezintă numai o parte din elementele unei colectivităţi sau cînd indicatorii sînt distribuiţi în unităţi diferite de timp, situate la intervale neegale (de exemplu, dacă s-ar prezenta producţia de oţel a României în 1950 ; 1970 ; 1978).

În practica statistică se folosesc uneori aşa-numitele grafice cu coloane în aflux. Ele se întîlnesc atunci cînd pentru fiecare grupă se foloseşte un „set" de coloane care se deplasează astfel încît să se vadă pe grafic diferenţele cantitative dintre subgrupe.



Scara: 1 cm=20%.
Legenda: ▨ depăşirea în % a planului.

Fig. 4.1. Gradul de îndeplinire a planului de producţie la întreprinderea *X* în luna ianuarie 1978.

Graficele prin benzi sînt construite tot în cadranul I din sistemul axelor rectangulare. Benzile se fixează pe axa ordonatelor, iar lungimea lor este proporţională cu valoarea indicatorilor de reprezentat. Deci, scara de reprezentare se stabileşte pe axa absciselor (*OX*) în raport cu valoarea maximă a indicatorilor. Benzile au, de regulă, o lăţime mult mai mică decît a coloanelor şi sînt separate prin distanţe egale. Acest tip de grafic se foloseşte în special pentru indicatori de lungime, sau cînd indicatorii prezintă între ei variaţii foarte mari, sau cînd datele sînt distribuite cronologic neegal. Tot o formă de grafic prin benzi este şi aşa-numita „piramidă a vîrstelor".

De exemplu, pentru a avea o imagine asupra dezvoltării producţiei de oţel în România, în anii socialismului, este suficient să se compare datele din tabelul 4.13.

Tabelul 4.13

Producţia de oţel a României în anii: 1950, 1970 şi 1977

Anii	Producţia de oţel, în mii tone
A	1
1950	555
1970	6 517
1977	11 457

Sursa: Anuarul Statistic al R.S. România, 1978, D.C.S., p. 162–169.

Scanned with OKEN Scanner

Dacă se consideră că 1 cm = 1 000, sînt necesare 12 diviziuni ale scării, iar datele se vor reprezenta ca în figura 4.2.

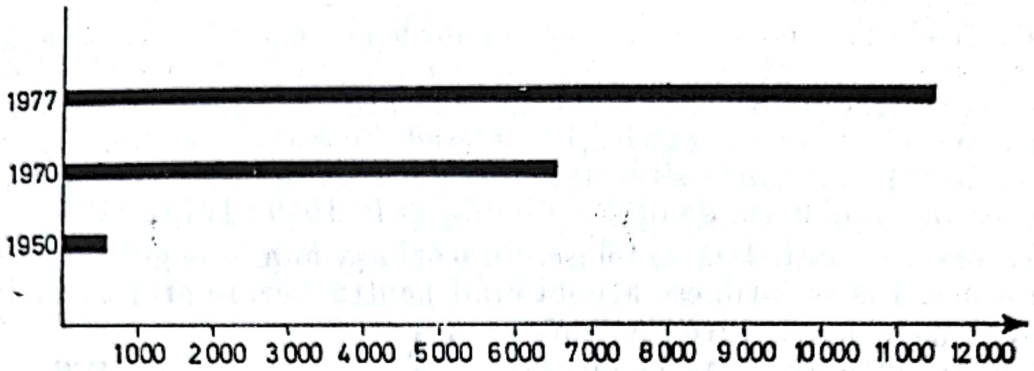


Fig. 4.2. Producția de oțel a R.S. România în anii 1950, 1970 și 1977.

Graficele prin figuri geometrice (*de suprafață sau de volum*) se folosesc fie pentru reprezentarea variației unor indicatori de volum, fie pentru reprezentarea structurii colectivității.

Diagramele de volum se aleg în raport cu conținutul indicatorilor de reprezentat. Dacă datele înregistrate se referă la fenomene independente, atunci se pot folosi figuri geometrice de suprafață unidimensionale (cerc, pătrat); dacă se înregistrează ca fenomene dependente de mai mulți factori, se folosesc figuri geometrice de suprafață sau volum (dreptunghi, paralelipiped) cu ajutorul cărora se reprezintă concomitent atît caracteristica complexă, cît și variația factorilor.

Diagrama prin pătrate se folosește atunci cînd se reprezintă grafic datele statistice centralizate la nivelul unei unități complexe, al unei grupe sau pe întreaga colectivitate.

De exemplu, se poate reprezenta prin diagramă prin pătrate numărul total al populației înregistrată la recensămintele din 1956, 1966 și 1977 în R.S. România.

Din Anuarul statistic al R.S. România pe anul 1977, pag. 45, se cunosc datele : 17 489 450 locuitori la 21 februarie 1956 ; 19 103 163 locuitori la 15 martie 1966 și 21 559 416 la 5 ianuarie 1977.

Considerînd că suprafețele celor două pătrate ar fi proporționale cu mărimea indicatorilor de reprezentat, latura pătratului se obține ca raport între rădăcina pătrată a valorii indicatorului și unitatea de lungime a scării de reprezentare a acestuia.

Se extrage rădăcina pătrată din valoarea indicatorilor pentru cei doi ani și, în funcție de spațiul avut la dispoziție, se alege o unitate de lungime a scării de reprezentare. Dacă se consideră 1 cm = 1 000 unități înseamnă că 1 cm^2 = 1 000 000 locuitori.

În interior se trec valorile reale ale indicatorilor care au fost considerați ca fiind proporționali cu suprafața pătratelor.

Pentru exemplul luat sînt necesare o serie de calcule.

Scanned with OKEN Scanner

Tabelul 4.14

Calcularea elementelor necesare reprezentării prin pătrate

Anii	Latura de reprezentat	Latura graficului 1 cm² = 1 mil. locuitori
0	1	2
1956	$\sqrt{17\,489\,450} \approx 4\,182$	$\frac{4\,182}{1\,000} \approx 4{,}2$ cm
1966	$\sqrt{19\,103\,163} \approx 4\,370$	$\frac{4\,370}{1\,000} \approx 4{,}4$ cm
1977	$\sqrt{21\,559\,416} \approx 4\,643$	$\frac{4\,643}{1\,000} \approx 4{,}6$ cm

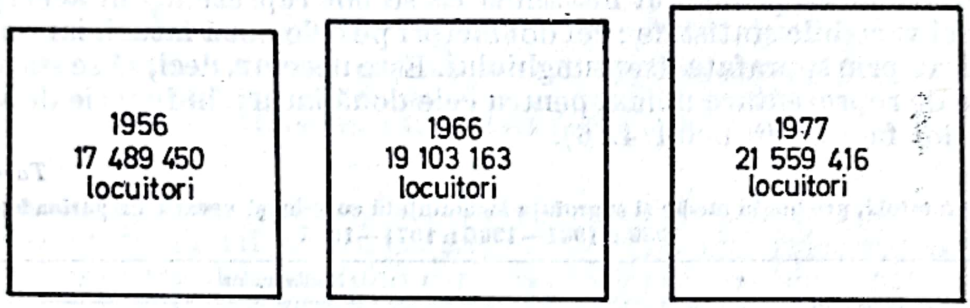


Fig. 4.3. Populația R.S. România în anii 1956, 1966 și 1977

Diagrama prin cercuri se folosește pentru aceleași cazuri ca și diagrama prin pătrate. La diagrama prin cercuri se consideră că valoarea indicatorilor este proporțională cu suprafața și unitatea de lungime a scării de reprezentare se fixează pe raza cercului.

Tabelul 4.15

Calcule necesare reprezentării prin cercuri

Anii	Raza de reprezentat	Raza cercului pe grafic 1 cm² = 1 mil. locuitori
A	1	2
1956	$\sqrt{\frac{17\,489\,450}{3{,}14}} \approx 2\,362$	$\frac{2\,362}{1\,000} \approx 2{,}4$ cm
1966	$\sqrt{\frac{19\,103\,163}{3{,}14}} \approx 2\,466$	$\frac{2\,466}{1\,000} \approx 2{,}5$ cm
1977	$\sqrt{\frac{21\,559\,416}{3{,}14}} \approx 2\,620$	$\frac{2\,620}{1\,000} \approx 2{,}6$ cm

Diagrama prin dreptunghiuri se folosește atunci cînd trebuie să se reprezinte nivelul totalizat al unei caracteristici complexe care se poate descompune în produsul a doi factori, de exemplu : prezentarea variației

Scanned with OKEN Scanner

fondului de retribuire în funcţie de retribuţia medie şi numărul de muncitori; sau a volumului producţiei în funcţie de nivelul de productivitate a muncii şi de numărul de muncitori etc.

Fig. 4.4. Populaţia R.S. România în anii 1956, 1966 şi 1977.

Folosind dreptunghiul înseamnă că se pot reprezenta în acelaşi timp cele trei variabile statistice : cei doi factori pe cele două laturi, iar variabila complexă prin suprafaţa dreptunghiului. Este necesar, deci, să se stabilească scările de reprezentare numai pentru cele două laturi, în funcţie de valorile celor doi factori (tabelul 4.16).

Tabelul 4.16

Producţia totală, producţia medie şi suprafaţa însămînţată cu grîu şi secară în perioadele 1934—1938; 1961—1965; 1971—1975

Indicatori	Media anilor		
	1934—1938	1961—1965	1971—1975
0	1	2	3
— suprafaţa cultivată (mii ha)	2 742,2	3 053,8	2 463,9
— producţia medie la ha (chintale)	10,2	14,5	22,9
— producţia totală (mii tone)	2 975,0	4 415,7	5 449,7

Sursa: Anuarul Statistic al R. S. România, 1977, D.C.S., p. 224—227.

Pentru a reprezenta grafic aceşti indicatori, se construiesc trei dreptunghiuri în care bazele vor fi proporţionale cu suprafaţa însămînţată, înălţimile cu producţia medie la ha, iar suprafeţele cu producţia totală obţinută (fig. 4.5.)

Se presupune că 1 cm = 3 q pe înălţime şi 1 cm = 1 mil. ha pe baza dreptunghiului; atunci cele trei dreptunghiuri ale graficului vor avea dimensiunile :

$$B_1 = \frac{2\,742{,}2}{1\,000} \approx 2{,}7 \text{ cm}; \quad B_2 = \frac{3\,053{,}8}{1\,000} \approx 3{,}1 \text{ cm};$$

$$B_3 = \frac{2\,463{,}9}{1\,000} \approx 2{,}5 \text{ cm};$$

$$I_1 = \frac{10{,}2}{3} \approx 3{,}4 \text{ cm}; \quad I_2 = \frac{14{,}5}{3} \approx 4{,}8 \text{ cm}; \quad I_3 = \frac{22{,}1}{3} \approx 7{,}4 \text{ cm}.$$

Scanned with OKEN Scanner

Stereograma presupune folosirea unui paralelipiped pentru a reda imaginea unei caracteristici statistice ce depinde de variaţia a trei factori înregistraţi.

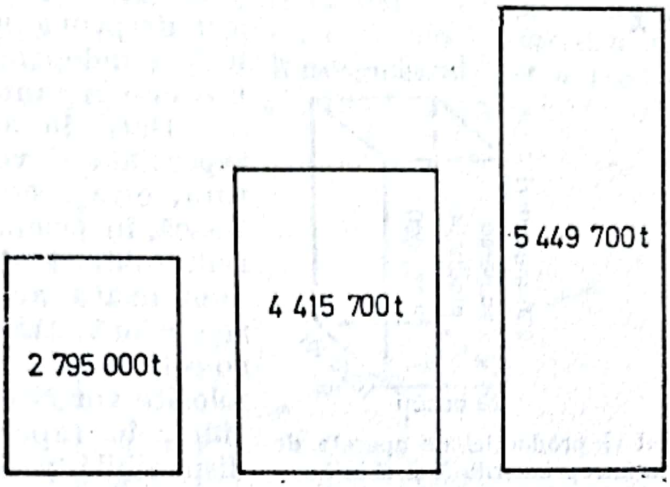


Fig. 4.5. Producţia totală medie şi suprafaţa însămînţată cu grîu şi secară în prioadel 1934—1938; 1961—1965; 1971—1975 (media anilor).

În cazul folosirii paralelipipedului este necesar să se stabilească scările de reprezentare numai pentru cei trei factori, deoarece produsul lor reprezintă volumul paralelipipedului care trebuie să fie proporţional cu mărimea caracteristicii complexe (vezi, de exemplu, datele din tabelul 4.17).

Tabelul 4.17

Costul total al aparatelor de radio de acelaşi tip, produse în întreprinderile X şi Z în funcţie de costul unitar, productivitatea muncii şi numărul mediu al muncitorilor

Întreprinderea	Numărul mediu scriptic al muncitorilor	Productivitatea medie lunară pe un muncit., în buc.	Costul unui aparat de radio, în lei	Cost total al producţiei lunare, în lei
A	1	2	3	4
X	300	30	1 650	14 850 000
Z	250	26	1 800	11 250 000

Sursa: D. Haşigan şi I. Marinescu, *Grafice şi elemente de calcul grafic, metode şi aplicaţii*, Ed. ştiinţifică Bucureşti, 1965, p. 133.

Dacă se consideră că înălţimea paralelipipedului este proporţională cu numărul muncitorilor (1 cm = 100 muncitori); lungimea bazei proporţională cu nivelul productivităţii muncii (1 cm = 20 bucăţi); lăţimea bazei proporţională cu costul de producţie (1 cm = 2 000 lei), atunci volumul paralelipipedului va reprezenta costul total al producţiei lunare şi se va înscrie pe grafic (fig. 4.6).

Diagramele de structură sînt folosite frecvent în prezentarea datelor statistice pentru interpretarea mutaţiilor intervenite în structura pe ra-

Scanned with OKEN Scanner

muri, sau pe plan teritorial. De asemenea, ele se pot folosi pentru orice colectivitate împărţită pe grupe, după variaţia uneia sau a mai multor caracteristici statistice. Graficele de structură se construiesc frecvent prin pătrate, cercuri, semicercuri sau dreptunghiuri. Ele se pot utiliza independent sau împreună cu diagramele de volum.

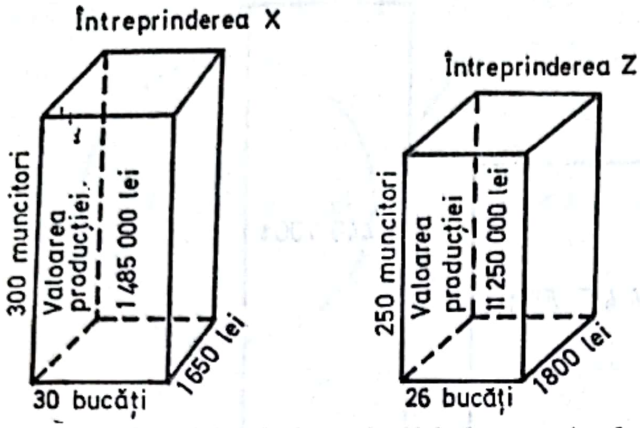


Fig. 4.6. Costul total al producţiei de aparate de radio de acelaşi tip la întreprinderile *X* şi *Z* în funcţie de costul unitar, productivitatea muncii şi numărul mediu al muncitorilor.

Dacă în acelaşi grafic se reprezintă şi volumul şi structura, este necesar să se stabilească, în primul rînd, scara de reprezentare după procedeele menţionate anterior. Dacă se reprezintă numai structura colectivităţii, atunci figurile folosite vor fi egale, ca dimensiuni, în raport cu suprafaţa disponibilă pentru grafic [1].

Cronograma (historiograma) se foloseşte în statistică pentru a reprezenta grafic o serie dinamică. Cronogramele se folosesc pentru a desprinde tendinţa de dezvoltare a fenomenelor pe fiecare etapă dată. Pentru aceasta se folosesc reţelele de grafice cunoscute, astfel încît să se poată reda cît mai fidel variaţia fenomenelor la care se referă datele.

Într-o serie dinamică valorile indicatorilor sînt reprezentate în funcţie de succesiunea lor în timp. Deci, variaţia de timp se va reprezenta pe axa absciselor, iar valoarea indicatorilor pe axa ordonatelor.

Dacă variaţia de timp este redată prin momente, atunci ele se trec în dreptul diviziunilor cotate pe axa absciselor. Dacă variaţia de timp este redată prin intervale de timp, atunci ele se trec între două diviziuni succesive. Indiferent de modul de redare a timpului pe axa absciselor, trebuie să se treacă toate unităţile succesive fără nici o întrerupere. Dacă în seria cronologică lipsesc date pentru unele valori de timp, atunci pentru perioada respectivă din interiorul seriei se face legătura printr-o linie întreruptă, considerînd tendinţa generală a întregii serii.

Pe axa ordonatelor se vor trece diviziunile scării astfel încît să cuprindă valoarea maximă a indicatorilor. De exemplu, trebuie să se reprezinte grafic datele din tabelul 4.18.

Tabelul 4.18

Producţia de energie electrică a R.S. România între anii 1970—1977

Anii	1970	1971	1972	1973	1974	1975	1976	1977
0	1	2	3	4	5	6	7	8
mil. Kwh	35 088	39 454	43 439	46 779	49 063	53 721	58 266	59 858

Sursa: Anuarul Statistic al R.S. România, D.C.S., 1978, p. 168—169.

[1] Modul de construire a graficelor de structură este prezentat în capitolul „Mărimi relative".

Scanned with OKEN Scanner

Fiind o serie cronologică de intervale, pe axa absciselor vor fi 8 diviziuni de timp, iar pe scara ordonatelor se va stabili unitatea de reprezentare, astfel încît să se asigure o proporţionalitate faţă de axa absciselor şi să cuprindă toate valorile indicatorilor. La cronogramă, în mod obligatoriu, scara ordonatelor începe cu zero, deci pentru a asigura proporţia necesară armoniei graficului în funcţie de valoarea maximă se poate lua 1 cm = 9 000 Kwh, obţinîndu-se astfel 7 diviziuni.

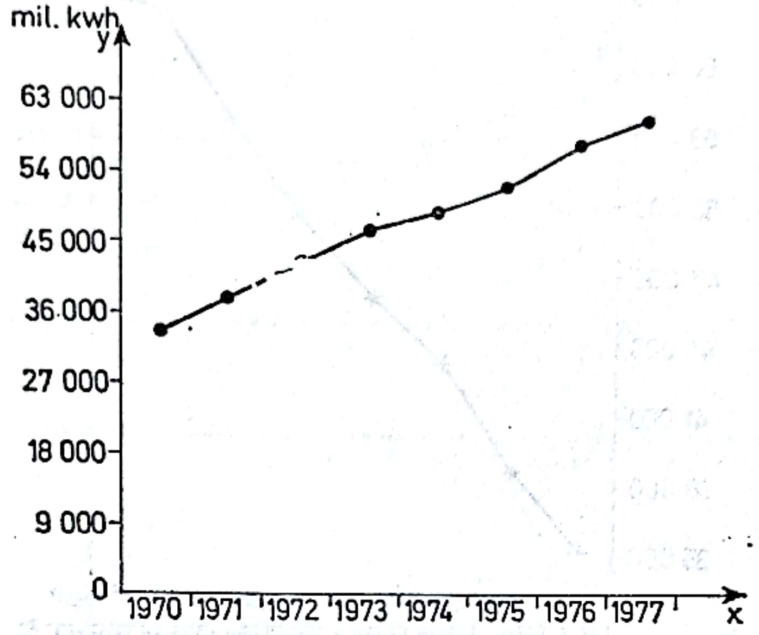


Fig. 4.7. Producţia de energie electrică în R.S. România între anii 1970—1977 (milioane Kwh).

Valoarea minimă fiind de 35 088, partea de jos a graficului rămîne neutilizată. De aceea, se recomandă să se facă o întrerupere de canal, adică să se aleagă scara de reprezentare nu în raport cu valoarea maximă, ci cu amplitudinea variaţiei A = 59 858 mil. — 35 088 mil. = 24 770 mil. Kwh.

Ţinînd seama de amplitudinea variaţiei şi păstrînd aceeaşi proporţionalitate între cele două axe se poate stabili o scară de reprezentare de 1 cm = 3 000 mil. Kwh. În noul grafic forma de variaţie care exprimă tendinţa de dezvoltare a producţiei de energie electrică apare şi mai clar (vezi fig. 4.8).

Seriile cronologice se pot reprezenta grafic şi pe scări neuniforme. Ele se reprezintă folosind scările logaritmice pe axa ordonatelor. Deoarece timpul se reprezintă pe scara uniformă, graficul se numeşte semilogaritmic. Acest grafic se utilizează, în special, atunci cînd apar fenomene ce prezintă variaţii foarte mari de la o perioadă la alta şi pentru completarea analizei se foloseşte scara logaritmică care reduce simţitor din variaţia termenilor seriei. De regulă, aceste grafice se folosesc ca mijloace complementare de analiză ; graficul de bază rămîne cel prezentat pe scara aritmetică.

Scanned with OKEN Scanner

La graficul semilogaritmic se procedează în felul următor :

— Se stabileşte înălţimea scării, care se divide în zece părţi egale (dreapta graficului). Aceeaşi scară se împarte apoi proporţional cu loga-

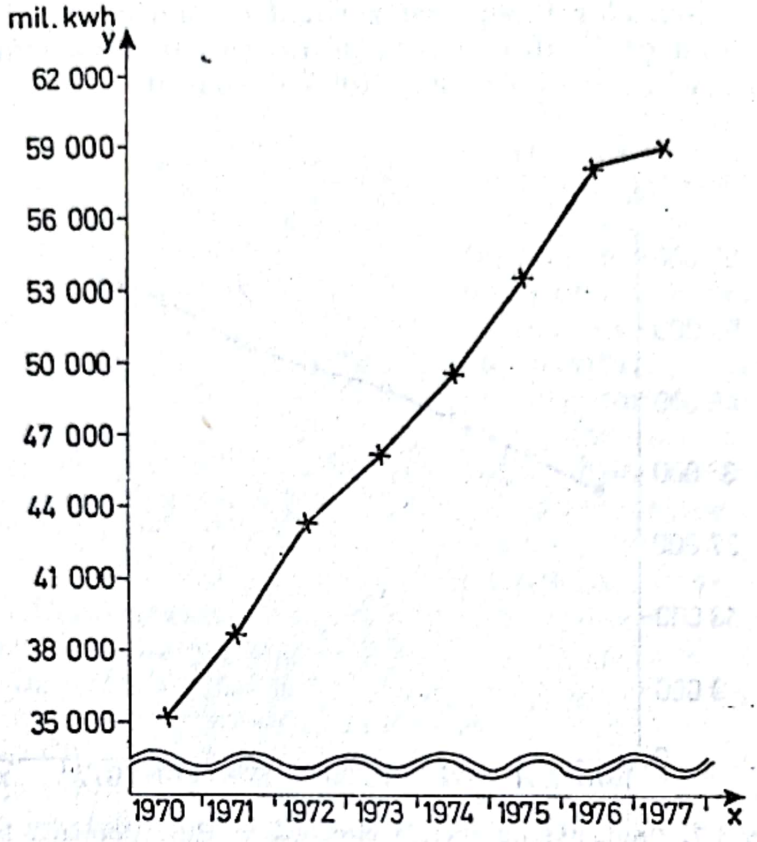


Fig. 4.8. Producţia de energie electrică în R.S. România între anii 1970—1977 (milioane Kwh).

ritmii numerelor 1—10. Dacă termenii seriei prin logaritmare au caracteristici diferite ca mărime, se folosesc atîtea module de cîte 10 părţi cîte sînt necesare diferenţei dintre caracteristici.

— Scara de reprezentare începe cu 1 şi nu cu 0 ca în cazul scării aritmetice, iar punctele cotate pe scara graficului se găsesc la distanţele corespunzătoare logaritmilor, deci scara este neuniformă.

Punctele de coordonate dintre variabila de timp şi variabila logaritmilor termenilor se unesc ca în cazul diagramelor cronologice pe scări uniforme.

— După ce se trasează curba (linia) se înlocuiesc logaritmii cu valorile indicatorilor absoluţi.

Pentru exemplificare se vor folosi aceleaşi valori ca şi în cazul cronogramei pe scară uniformă (vezi fig. 4.9).

O altă problemă care trebuie analizată la seriile cronologice este şi aceea a sezonalităţii. În statistica social-economică se întîlnesc cazuri în care schimbarea anotimpurilor provoacă variaţii mai mari sau mai mici a nivelului de dezvoltare a fenomenelor şi de la o lună la alta, sau de la un trimestru la altul, de exemplu, consumul casnic de gaze naturale, cantitatea producţiei de bere, volumul turismului etc.

Scanned with OKEN Scanner

Pentru a putea interpreta gradul şi forma de variaţie sezonieră, se foloseşte diagrama polară sau radială. Poartă această denumire deoarece la acest grafic se foloseşte o reţea polară formată din cercuri concentrice, iar

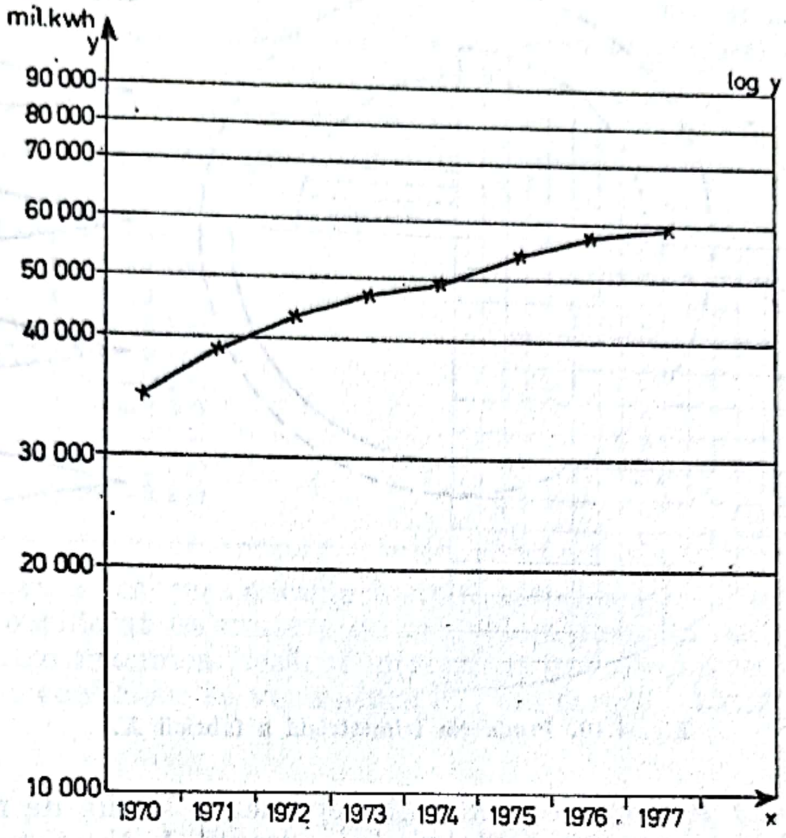


Fig. 4.9. Producţia de energie electrică în R.S. România între anii 1970—1977 (milioane Kwh).

raza este proporţională cu nivelul mediu al indicatorilor. Cercul se împarte în atîtea sectoare cîţi indicatori sînt. Dacă variaţia este lunară, atunci se obţin 12 sectoare de cerc, care sînt mărginite de cîte două raze proporţionale cu mărimea indicatorilor de reprezentat.

Dacă valoarea unui indicator depăşeşte media valorilor individuale, atunci se vor prelungi cele două raze în afara cercului, iar dacă valoarea lor este mai mică, atunci se vor îngroşa razele numai pînă la punctele corespunzătoare pe cele două raze.

După ce s-au fixat toate punctele pe reţeaua polară, se unesc aceste puncte prin linii frînte sau curbe. Comparînd cu linia cercului de bază, se poate interpreta cît este de mare variaţia fiecărei luni sau fiecărui trimestru în raport cu valoarea care ar fi trebuit să fie realizată dacă fenomenul n-ar fi fost influenţat de valul sezonier.

Tabelul 4.19

Producţia trimestrială a fabricii

(date convenţionale)

Trimestrul	I	II	III	IV	Media trimestrială
Producţia (mii kg)	2 500	3 500	7 000	3 000	4 000

Scanned with OKEN Scanner

Dacă se consideră scara de reprezentare 1 cm = 100 mii kg, se va folosi un cerc de bază cu o rază de 4 cm (vezi fig. 4.10).

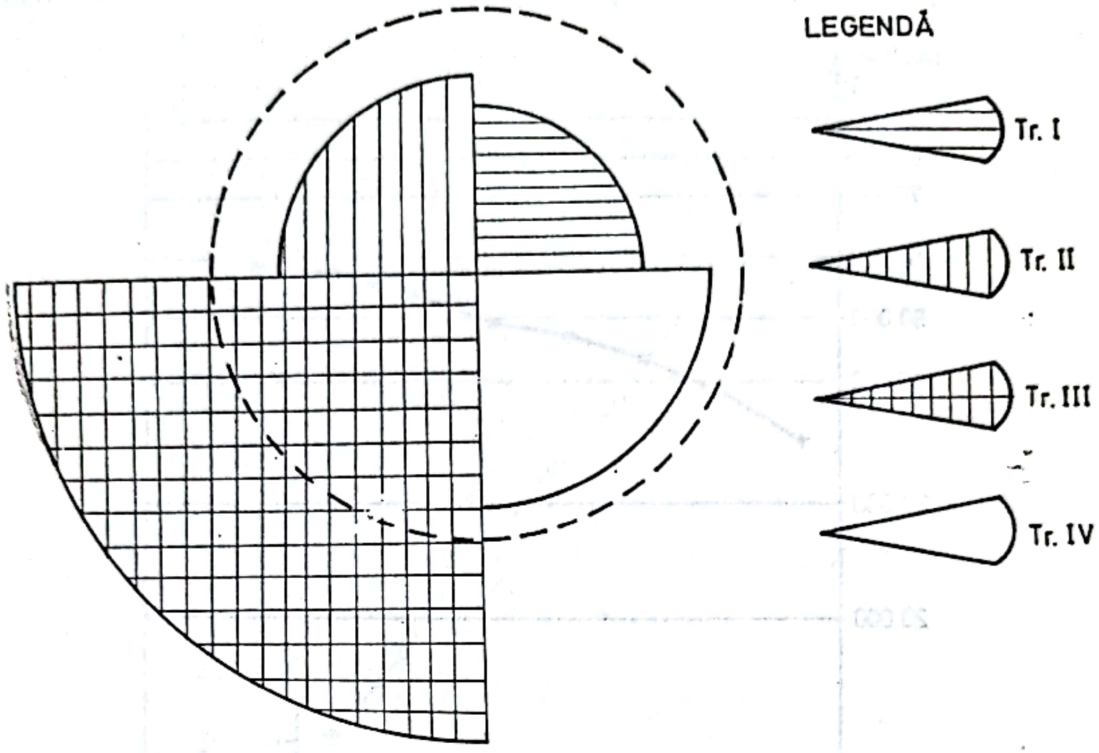


Fig. 4.10. Producţia trimestrială a fabricii *X*.

Histograma se foloseşte pentru a reprezenta seriile de repartiţie de frecvenţe. Pentru aceasta se foloseşte cadranul I din sistemul de axe rectangulare; pe axa absciselor se trec valorile caracteristicii de grupare, iar pe axa ordonatelor se stabileşte scara pentru reprezentarea frecvenţelor. La construirea histogramei trebuie să se ţină seama de felul intervalelor de grupare. Dacă intervalele sînt egale, pe axa absciselor se trec valorile care marchează limitele succesive ale intervalelor şi apoi se stabileşte scara frecvenţelor pe ordonată în funcţie de mărimea frecvenţei maxime. Grupele se reprezintă prin coloane a căror suprafaţă trebuie să fie egală cu produsul dintre mărimea intervalului şi frecvenţă. Deasupra lor, de obicei, se trec frecvenţele corespunzătoare fiecărei grupe.

De exemplu, se poate reprezenta grafic, printr-o histogramă seria de repartiţie ca în tabelul 4.20.

În mod convenţional, pe scara absciselor, la histogramă primul punct cotat corespunde cu valoarea minimă a caracteristicii de grupare, adică pentru seria luată ca exemplu primul punct cotat pe scara absciselor va fi 1 000, iar ultimul 3 000; avînd zece intervale se vor obţine zece diviziuni egale ale scării. Pentru scara ordonatelor valoarea maximă fiind 70, se poate lua 1 cm = 10 muncitori (vezi fig. 4.11).

Interpretînd graficul, se poate observa că la această serie există tendinţă de concentrare către valorile centrale ale retribuţiilor primite, deci seria prezintă o uşoară asimetrie, valorile mai mari fiind mai bine reprezentate în comparaţie cu cele mai mici ale caracteristicii înregistrate.

Scanned with OKEN Scanner

Dacă intervalele sînt neegale între ele, histograma se construieşte respectînd principiul de bază potrivit căruia suprafaţa unei coloane trebuie să fie proporţională cu produsul dintre mărimea intervalului şi frec-

Tabelul 4.20

Repartizarea muncitorilor întreprinderii A după mărimea retribuţiei nete din luna ianuarie 1978

(date convenţionale)

Grupe de muncitori după mărimea retribuţiei	Numărul muncitorilor
sub 1 600	16
1 600—1 800	25
1 800—2 000	45
2 000—2 200	53
2 200—2 400	70
2 400—2 600	63
2 600—2 800	52
2 800—3 000	41
3 000—3 200	25
3 200—3 400	10

venţa sa. Pentru a respecta acest principiu, este necesar să se stabilească un interval considerat ca unitate de măsură şi să se măsoare proporţia care există între mărimea fiecărui interval şi cea a intervalului luat ca bază. Cu acest coeficient se vor reduce în mod corespunzător frecvenţele

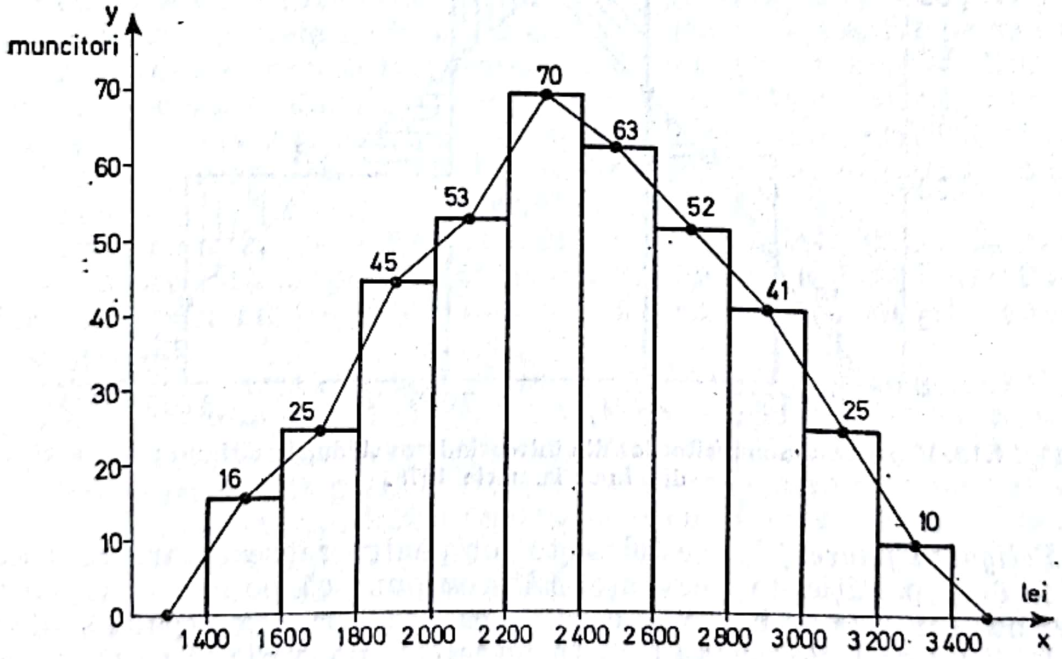


Fig. 4.11. Repartizarea muncitorilor din întreprinderea *A* după mărimea retribuţiei nete din luna ianuarie 1978

reale, pentru a obţine frecvenţele reduse, în funcţie de care se va stabili scara de reprezentare pe axa ordonatelor.

Dacă aceeaşi colectivitate de muncitori se regrupează pe intervale în 4 grupe, se obţine situaţia din tabelul 4.21.

Scanned with OKEN Scanner

Tabelul 4.21

Repartizarea muncitorilor din întreprinderea A după mărimea retribuției din luna ianuarie 1978

Grupe de muncitori după retribuția netă primită	Numărul muncitorilor (f_i)	Mărimea intervalului (k_i)	Coeficientul de reducere a frecvențelor $\frac{k_i}{k_0}$	Frecvențe reduse $f_i : \frac{k_i}{k_0}$
A	1	2	3	4
1 400—1 600	16	200	1	16
1 600—2 000	70	400	2	35
2 000—2 600	186	600	3	62
2 600—3 400	128	800	4	32

Pe axa absciselor se vor reprezenta cele zece unități de interval, dar se vor marca numai punctele care corespund limitelor intervalului. Pe axa ordonatelor se va stabili scara de reprezentare a frecvențelor care să cuprindă valoarea maximă a frecvențelor reduse, iar deasupra coloanelor se vor trece valorile reale ale acestora.

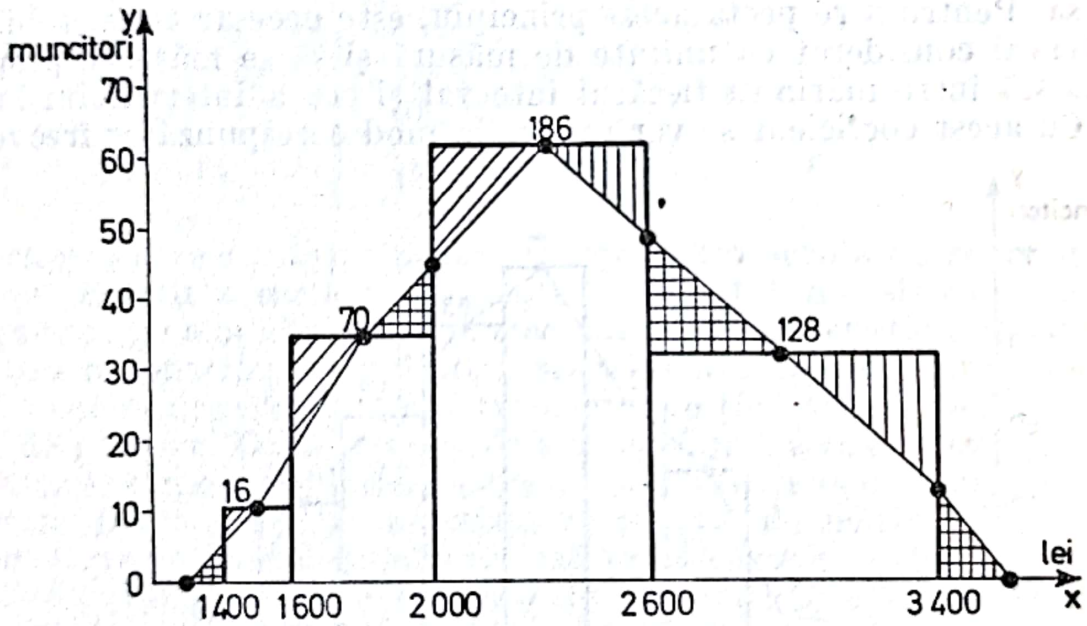


Fig. 4.12. Repartizarea muncitorilor din întreprinderea *A* după mărimea retribuției din luna ianuarie 1978.

Poligonul frecvențelor se folosește tot pentru reprezentarea grafică a seriilor de repartiție de frecvențe. El presupune că pe axa absciselor se reprezintă centrele de interval, care, în cazul intervalelor egale, se distribuie la distanțe egale pe axa *OX*. În acest caz, pe grafic se va reprezenta linia care unește punctele de coordonate dintre valorile centrelor de interval și frecvența lor. Dacă poligonul frecvențelor se construiește pe baza histogramei, atunci el se obține prin unirea centrelor din coloanele histogramei (vezi fig. 4.11). Se poate observa că pe grafic linia care corespunde relației dintre valorile caracteristicii și frecvențele lor delimitează aceeași arie ca și la histogramă.

Scanned with OKEN Scanner

Dacă intervalele de grupare sînt neegale, atunci se procedează ca şi în cazul histogramei: se reduc mai întîi proporţional frecvenţele şi apoi se reprezintă în funcţie de mărimea intervalului (vezi fig. 4.12).

Curba cumulativă a frecvenţelor se foloseşte atunci cînd se determină pe grafic valorile mediilor de poziţie (mediana, cuartilele, decilele, percentilele)[1]. Curba cumulativă a frecvenţelor presupune însumarea succesivă a frecvenţelor în ambele sensuri ale seriei (tabelul 4.22).

Tabelul 4.22

Repartizarea muncitorilor din întreprinderea A după mărimea retribuţiei din luna ianuarie 1978

Grupe de muncitori după retribuţia lunară	Numărul muncitorilor	Frecvenţe cumulate	
		sub limita superioară	sub limita inferioară
A	1	2	3
sub 1 600	16	16	400
1 600—1 800	25	41	384
1 800—2 000	45	86	359
2 000—2 200	53	139	314
2 200—2 400	70	209	261
2 400—2 600	63	272	191
2 600—2 800	52	324	128
2 800—3 000	41	365	76
3 000—3 200	25	390	35
3 200—3 400	10	400	10

Pentru a construi curba cumulativă a frecvenţelor, se reprezintă pe axa absciselor centrele de interval echidistante între ele, iar pe axa ordonatelor se vor reprezenta frecvenţele cumulate într-un sens şi altul. Deci, pe grafic vor apare două curbe care unesc punctele de coordonate dintre centrele de interval şi frecvenţele cumulate, în ordine crescătoare şi descrescătoare. Punctul de intersecţie a celor două curbe va marca tendinţele de asimetrie a seriei prezentate grafic (Fig. 4.13).

Curba cumulativă a frecvenţelor se mai numeşte şi ogivă, în cazul cînd seria prezintă un pronunţat accent de simetrie a distribuţiei frecvenţelor, în raport cu frecvenţa maximă, ce corespunde valorii centrale a caracteristicii.

Pentru seria luată ca exemplu, pe axa absciselor se vor marca diviziunile scării cuprinse între 1 400 şi 3 400 (1 cm = 200 lei), iar pe scara ordonatelor diviziunile scării cuprinse între 10 şi 400 (1 cm = 400 muncitori).

Cartogramele sînt graficele care se folosesc pentru prezentarea densităţii de repartiţie teritorială a fenomenelor cu ajutorul hărţilor. De exemplu, dacă se prezintă grafic gradul de finisare a lucrărilor agricole de toamnă pe judeţele ţării, atunci se va folosi o hartă a ţării în care densitatea de repartiţie se va reprezenta fie prin intensităţi de culoare, fie prin intensităţi de haşuri, iar în interiorul fiecărui judeţ se vor trece valorile înregistrate ale caracteristicii a cărei variaţie teritorială se prezintă. Citeodată se ierarhizează valorile caracteristicii şi numerele de ordine sau rangurile, cum se mai numesc în statistică, se înscriu pe hartă.

[1] Aceşti indicatori se vor trata la capitolul „Mărimile medii, indicatorii variaţiei şi asimetriei".

Scanned with OKEN Scanner

Cartodiagramele presupun localizarea pe hartă a diferitelor diagrame statistice. Deci, pe hartă se vor construi diferite figuri geometrice (pătrate, cercuri, dreptunghiuri etc.), pentru a reda volumul sau structura diferiţilor indicatori distribuiţi din punct de vedere teritorial.

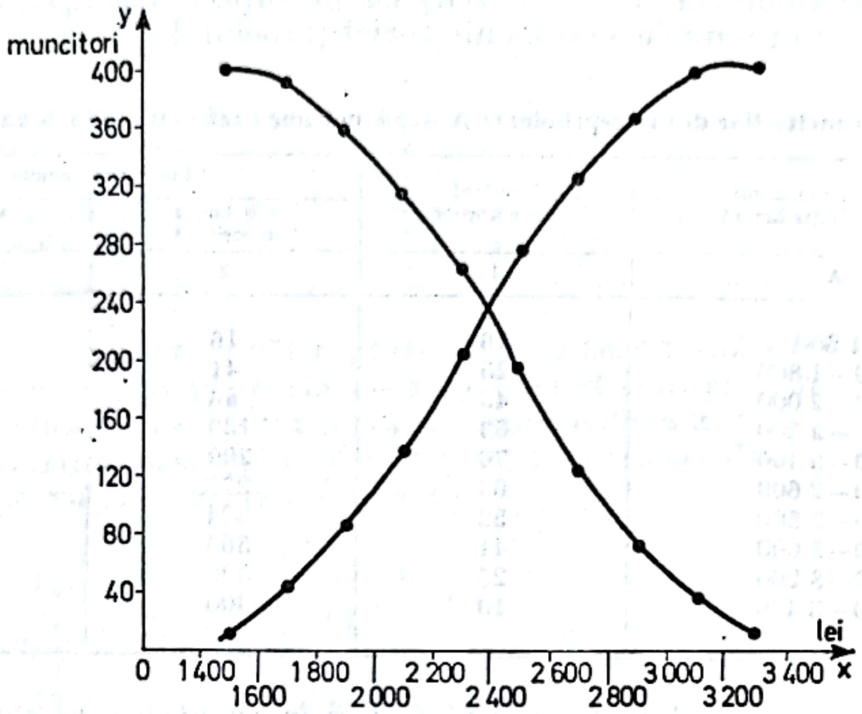


Fig. 4.13. Repartizarea muncitorilor din întreprinderea *A* după mărimea retribuţiei din luna ianuarie 1978.

La reprezentarea grafică a distribuţiilor teritoriale ale diferiţilor indicatori se mai pot folosi şi figuri naturale sau simbolice, care sînt proporţionale cu valoarea indicatorilor de reprezentat.

De remarcat că figurile naturale şi simbolice se folosesc în practica statistică nu numai pentru repartiţia teritorială, ci ori de cîte ori un astfel de grafic este mai sugestiv şi ajută la popularizarea datelor statistice.

În concluzie, formele de reprezentare grafică în statistica social-economică sînt extrem de numeroase, ceea ce impune alegerea cu discernămînt a acelora care sînt mai adecvate conţinutului indicatorilor şi pot sugera cu uşurinţă tendinţele de manifestare a fenomenelor studiate. Reprezentările grafice pot fi folosite în statistică pentru prezentarea rezultatelor sau ca mijloc de analiză a formelor de manifestare a fenomenelor în vederea aplicării diferenţiate a metodelor de calcul statistic. În funcţie de scopul urmărit, se va alege graficul cel mai adecvat şi care să respecte riguros principiile de construire corectă a graficului. Deosebit de importante pentru aceasta sînt alegerea scărilor de reprezentare şi asigurarea proporţionalităţii graficului în funcţie de dimensiunile indicatorilor şi spaţiului de care se dispune pentru figura graficului. Dacă se reprezintă grafic indicatori cu valori mari şi se foloseşte cadranul I din sistemul axelor rectangulare, atunci se pot face întreruperi de canal pentru a reda mai bine forma de variaţie a caracteristicilor.

Scanned with OKEN Scanner

CAPITOLUL V

MĂRIMI RELATIVE

În procesul de cercetare statistică se obţin frecvent indicatori care prezintă, pentru aceeaşi caracteristică, diferenţe de nivel de la o unitate la alta, de la o grupă la alta sau de la o colectivitate la alta. Pentru a le supune analizei statistice în vederea măsurării şi explicării acestor diferenţe de nivel este necesar să se treacă la compararea acestor indicatori. De cele mai multe ori rezultatul comparării în statistica social-economică se exprimă cu ajutorul mărimilor relative.

5.1. Noţiunea de mărime relativă în statistica social-economică

O primă categorie de indicatori derivaţi, cu largă răspîndire în statistică, o formează mărimile relative care sintetizează, în expresii numerice, rezultatele comparării datelor statistice în interiorul aceleiaşi colectivităţi de la o grupă la alta sau de la o colectivitate la alta. Rezultatul comparării sub formă de raport a doi indicatori statistici este o *mărime relativă*, care arată cîte unităţi din indicatorul de raportat revin la o unitate a indicatorului bază de raportare.

Mărimile relative sînt folosite în toate domeniile în care se aplică metodele de măsurare şi analiză statistică. Ele au proprietatea de a sugera, printr-o expresie numerică, proporţiile dintre indicatorii supuşi comparării.

Modul de calcul al mărimilor relative este simplu : ele se obţin prin efectuarea raportului dintre cei doi indicatori care trebuie să fie comparaţi.

Problemele care se pun în legătură cu folosirea mărimilor relative se referă la alegerea bazei de comparare, asigurarea comparabilităţii datelor care formează raportul, alegerea formei de exprimare a mărimii relative.

Alegerea bazei de comparare se face în funcţie de gradul de interdependenţă dintre caracteristicile sau fenomenele comparate sau în funcţie de scopul cercetării. Cînd comparaţia se face în dinamică, alegerea bazei de raportare este relativ simplă, deoarece se ia ca bază de referinţă fie o perioadă sau un moment considerat ca fiind semnificativ pentru dezvoltarea fenomenelor pe întreaga perioadă, fie o perioadă de timp (moment) precedentă.

Probleme mai dificile apar însă cînd comparaţiile se fac pe plan teritorial. În acest caz alegerea bazei de raportare trebuie să se facă prin analiza fiecărei perioade în parte, în funcţie de modificările teritoriale ale diferitelor caracteristici sociale şi economice care au, de regulă, ritmuri foarte diferenţiate de dezvoltare.

Alegerea bazei de raportare trebuie să se facă astfel încît rezultatul raportului să fie cît mai semnificativ, corespunzător unor relaţii reale între

Scanned with OKEN Scanner

fenomenele supuse comparației, mai ales că valoarea raportului depinde, în principal, de mărimea din numitorul fracției. De exemplu, dacă se stabilește o serie de mărimi relative calculate față de cel mai mic nivel al seriei, aceste mărimi vor avea valori mult mai mari decît dacă s-ar raporta toate nivelurile individuale sau parțiale față de valoarea lor centrală. Dacă se ține seama că în cadrul fenomenelor de masă se manifestă tendința de concentrare a unităților către valoarea lor centrală, atunci este evident faptul că rapoartele calculate față de această mărime au o putere mai mare de sintetizare.

Asigurarea comparabilității datelor care formează raportul este problema esențială care trebuie rezolvată înainte de a trece la calcularea mărimilor relative. Aceasta înseamnă, de fapt, ca cei doi indicatori raportați să fie mărimi comparabile între ele. De exemplu, dacă se calculează valoarea producției nete ce revine la 1 000 lei fonduri fixe, atunci este necesar ca valoarea fondurilor fixe să se ia ca medie anuală pentru a corespunde cu valoarea producției nete care se calculează pe aceeași perioadă.

De asemenea, dacă în perioada analizată au intervenit modificări organizatorice — comasări de întreprinderi etc. —, atunci, înainte de a trece la calcularea mărimilor relative, trebuie reconstituită seria de date statistice comparabile.

Și din acest punct de vedere, mărimile relative calculate pe plan teritorial și, în special, cele care se referă la comparațiile teritoriale între țări, ridică cele mai multe probleme. În această situație, se pot supune calculului comparativ numai acele mărimi care corespund ca metodologie de culegere, prelucrare și grad de cuprindere ale elementelor ce definesc conținutul indicatorilor la care se referă.

Alegerea formei de exprimare a mărimilor relative se face pornind de la diferențele de mărime care există între indicatorii comparabili. Pentru a face sugestivă o mărime relativă, în statistică este necesar să se aleagă în mod corespunzător forma sa de exprimare. Mărimile relative pot fi exprimate sub formă de : numere întregi, fracții ordinare și zecimale, procente, promile, prodecimile etc.

Exprimarea sub formă de numere întregi, fracții zecimale sau ordinare se face atunci cînd cei doi indicatori comparați dau un cît exact sau sînt valori apropiate. Această formă de exprimare se mai numește în mod curent și coeficient, deoarece el se obține direct din rezolvarea raportului. În cazul în care exprimarea sub formă de coeficient nu este destul de sugestivă, deoarece indicatorul de raportat este mult prea mic față de indicatorul ales ca bază de raportare, se mărește proporțional valoarea raportului de 100, 1 000, 10 000 ori etc. De exemplu : într-o secție, comparînd producția lunii februarie cu producția lunii ianuarie, se obține raportul :

$$K = \frac{112\,500}{100\,000} = 1{,}125 \text{ ori.}$$

Scanned with OKEN Scanner

Mai sugestiv devine acest coeficient dacă se exprimă în procente, adică se arată cîţi lei din producţia lunii februarie revin la 100 lei din producţia lunii ianuarie :

$$K = \frac{112\,500}{100\,000} \cdot 100 = 1{,}125 \cdot 100 = 112{,}5\%.$$

Exprimarea sub formă de promile (‰) se foloseşte în special în demografie la calculul coeficienţilor mişcării naturale şi migratorie a populaţiei. De exemplu, coeficientul de natalitate arată care este proporţia de născuţi vii la 1 000 de locuitori şi se obţine făcînd raportul :

$$K = \frac{\text{numărul născuţilor vii}}{\text{numărul mediu al populaţiei}} \cdot 1\,000 = x‰. \qquad (5.1)$$

În cazul exprimării sub formă de prodecimile (‱) se arată cîte unităţi din indicatorul de raportat revin la 10 000 unităţi din indicatorul bazei de raportare, de exemplu :

$$K = \frac{\text{numărul studenţilor dintr-o ţară}}{\text{numărul populaţiei din ţara respectivă}} \cdot 10\,000 = y‱. \qquad (5.2)$$

În concluzie, alegerea formei de exprimare a mărimilor relative depinde de raportul de mărime dintre cele două valori comparate, astfel încît ele să sugereze cu claritate proporţia reală dintre datele supuse comparaţiei.

5.2. FELURILE MĂRIMILOR RELATIVE

Larga utilizare a mărimilor relative în practica statistică impune o clasificare a lor în funcţie de domeniul în care se aplică şi scopul în care se calculează. În rîndul lor deosebim : mărimi relative ale planului, mărimi relative ale structurii, mărimi relative de coordonare, mărimi relative ale dinamicii, mărimi relative de intensitate.

Mărimile relative ale planului sînt specifice statisticii din ţările socialiste, ale căror economii naţionale se dezvoltă pe bază de plan. Statistica are sarcina de a măsura gradul de îndeplinire a planului, analizînd influenţa diferiţilor factori asupra modului de realizare a sarcinilor de plan, scoţînd în evidenţă rezervele neutilizate.

Mărimile relative ale planului sînt de două feluri : mărimi relative ale sarcinii de plan şi mărimi relative ale îndeplinirii planului.

Mărimile relative ale sarcinii de plan arată măsura în care se preconizează să se depăşească nivelul care a fost atins de un fenomen în perioada

Scanned with OKEN Scanner

aleasă ca bază de raportare (de obicei, perioada precedentă). Dacă se notează volumul planificat cu x_{pl} şi cel realizat în perioada de bază cu x_0, atunci acest coeficient se calculează după formula :

$$K_{pl/0} = \frac{x_{pl}}{x_0} \cdot 100, \qquad (5.3)$$

în care :

$K_{pl/0}$ reprezintă mărimea relativă a sarcinii de plan (fracţia pl/0 arătînd tocmai între ce perioade se face comparaţia).

De exemplu, într-o întreprindere producţia realizată în luna decembrie a anului precedent a fost de 5 000 bucăţi cu un cost efectiv de 30 lei pe bucată. În luna ianuarie a anului curent, ţinînd seama de condiţiile de producţie existente, întreprinderea îşi propune să producă 5 300 bucăţi cu un cost de 29 lei pe bucată. Mărimile relative ale sarcinii de plan vor fi :

$$K_{pl/0} = \frac{x_{pl}}{x_0} \cdot 100 \begin{cases} \text{pentru producţie } K_{pl/0} = \dfrac{5\,300}{5\,000} \cdot 100 = 106\,\% \\ \text{pentru costul pe produs } K_{pl/0} = \dfrac{29}{30} \cdot 100 = 96{,}67\,\% \end{cases}$$

Producţia va trebui să crească în luna ianuarie a anului curent faţă de producţia lunii decembrie a anului precedent cu cel puţin 6 %, în timp ce costul va trebui să scadă cel puţin cu 3,33 %.

Deci, în primul caz este o mărime relativă supraunitară, iar în al doilea caz este o mărime relativă subunitară. De aceea interpretarea mărimilor relative ale sarcinii de plan trebuie să se facă după natura fenomenelor la care se referă. Astfel, pentru fenomene care au tendinţă obiectivă de creştere se vor stabili sarcini de plan minime, adică producţia va trebui să crească cel puţin la 106 %, în timp ce pentru fenomenele care au tendinţă de scădere se vor da sarcini de plan maxime, respectiv costul trebuie să ajungă cel mult la 96,67 % în comparaţie cu perioada de bază. De reţinut este faptul că sarcinile de plan minime şi maxime nu trebuie confundate cu limitele minime sau maxime care se stabilesc în cazul planurilor de perspectivă.

Mărimile relative ale îndeplinirii planului măsoară gradul de realizare a sarcinilor prevăzute în plan. Dacă se notează cu x_1 nivelul realizat, atunci coeficientul îndeplinirii planului se calculează după formula :

$$K_{1/pl} = \frac{x_1}{x_{pl}}. \qquad (5.4)$$

Continuînd exemplul anterior se presupune că producţia realizată în luna ianuarie a anului trecut a fost de 5 700 bucăţi, iar costul pe bucată

Scanned with OKEN Scanner

de 27 lei, atunci mărimile relative ale îndeplinirii planului la cei doi indicatori vor fi :

$$K = \frac{x_1}{x_{pl}} \begin{cases} \text{pentru producție } K_{1/pl} = \dfrac{5\,700}{5\,300} \cdot 100 = 107{,}5\% \\ \text{pentru cost} \qquad K_{1/pl} = \dfrac{27}{29} \cdot 100 = 93{,}1\% \end{cases}$$

Interpretarea gradului de îndeplinire și depășire a planului se face în funcție de natura sarcinii de plan. În cazul sarcinii de plan minime se consideră depășire a planului tot ceea ce depășește 100%, deci ea va fi cu atît mai mare cu cît nivelul realizat al fenomenului este mai mare decît nivelul său planificat. Cînd există un fenomen pentru care s-a prevăzut o sarcină maximă, depășirea se calculează tot față de 100%, care reprezintă limita maximă, deci depășirea se va obține făcînd diferența dintre mărimea realizată a îndeplinirii planului și 100%, obținînd o valoare negativă corespunzătoare unei scăderi mai mari a nivelului fenomenului decît cel prevăzut în plan. În acest caz, cu cît nivelul realizat este mai mic în raport cu nivelul planificat cu atît sarcina de plan este depășită în mai mare măsură. Rezultă că sarcinilor minime le corespund procente de depășire cu semn pozitiv, iar sarcinilor de plan maxime, procente de depășire a sarcinii planificate cu semn negativ.

Mărimea relativă de depășire a planului se calculează ca diferență între mărimea relativă de îndeplinire a planului și 100%, adică :

$$\text{Depășire relativă a planului } \left(\frac{x_1}{x_{pl}} 100\right) - 100. \tag{5.5}$$

Deci în exemplul considerat, producția planificată a fost depășită cu 7,5% (107,5% — 100 = 7,5%) iar costul s-a redus față de cel planificat cu 6,9% (93,1% — 100% = —6,9%).

În documentele de partid se întîlnesc numeroase exemple de astfel de mărimi relative. Este suficient să citățm cîteva exemple din planul 1976—1980 (tab. 5.1).

Mărimile relative de structură și reprezentarea lor grafică. Frecvent, colectivitățile supuse cercetării se împart în grupe omogene după variația uneia sau mai multor caracteristici statistice. Pentru a stabili care este raportul dintre fiecare grupă și întregul ansamblu este necesar să se calculeze mărimile relative de structură.

Mărimile relative de structură (ponderi sau greutăți specifice) se pot calcula atît pentru frecvențele grupelor, cît și pentru valorile centralizate ale diferitelor caracteristici.

Frecvențele relative se obțin raportînd frecvența absolută a fiecărei grupe la suma frecvențelor pe întreaga colectivitate, după formula :

$$f_i^* = \frac{f_i}{\sum_{i=1}^{m} f_i} \tag{5.6}$$

Scanned with OKEN Scanner

în care:

f_i^* reprezintă frecvența relativă sau greutatea specifică sau ponderea fiecărei grupe față de totalul frecvențelor;

f_i — frecvența absolută, obținută prin centralizarea pe grupe a unităților colectivității;

Σ — semnul sumei.

Tabelul 5.1[1]

Ritmurile medii de creștere la principalii indicatori, prevăzute în planul cincinal 1976 — 1980 (În procente, calculate pe baza prețurilor la 1.1.1977)

Indicatori	Ritm mediu anual 1976—1980	
	Plan cincinal	Program suplimentar
A	1	2
Producția globală industrială	10,7	11,5
Producția globală agricolă	6,5—8,6	6,9—9,0
Volumul investițiilor în economia națională (pe întreaga perioadă de cinci ani)	10,3*	12,8
Desfacerile de mărfuri prin comerțul socialist	8,1	8,7
Prestările de servicii către populație	11,0	11,9
Productivitatea muncii în industria socialistă	8,5	9,2
Venitul național	10,3	11,0

* Planul cincinal actualizat ca urmare a acțiunii de reducere a costurilor investițiilor.

Mărimi relative de structură se pot obține pentru toate caracteristicile care au fost centralizate pe grupe. În acest caz se raportează nivelul totalizat al caracteristicii din fiecare grupă (x_i) la valoarea centralizată pe total colectivitate $\left(\sum_{i=1}^{m} x_i\right)$ adică:

$$x_i^* = \frac{x_i}{\sum_{i=1}^{m} x_i}. \qquad (5.7)$$

Mărimile relative de structură prezintă următoarea proprietate: dacă se calculează toate mărimile relative de structură corespunzătoare tuturor grupelor în care s-a împărțit colectivitatea după variația aceleiași caracteristici și se face suma lor, aceasta trebuie să fie egală cu 1, sau dacă aceste greutăți specifice sînt exprimate în procente, cu 100, adică:

$$\frac{f_1}{\sum_{i=1}^{m} f_i} + \frac{f_2}{\sum_{i=1}^{m} f_i} + \cdots + \frac{f_m}{\sum_{i=1}^{m} f_i} = \frac{\sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} f_i} = 1,$$

[1] Nicolae Ceaușescu, *Raport la Conferința Națională a Partidului Comunist Român*, 7—9 decembrie 1977, Editura politică, București, 1977, p. 16.

Scanned with OKEN Scanner

sau în procente :

$$\left(\frac{f_1}{\sum_{i=1}^{m} f_i}\cdot 100\right)+\left(\frac{f_2}{\sum_{i=1}^{m} f_i}\cdot 100\right)+\cdots\cdots+\left(\frac{f_m}{\sum_{i=1}^{m} f_i}\cdot 100\right)=$$

$$=\left(\frac{\sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} f_i}\cdot 100\right)=100. \qquad (5.8)$$

Deci :

$$\sum_{i=1}^{m} f_i^{*}\% = 100$$

Aceeaşi relaţie este valabilă şi pentru greutăţile specifice calculate pentru variaţia pe grupe a unei caracteristici însumabile direct $\left(\frac{x_i}{\sum_{i=1}^{m} x_i}\right)$.

Mărimile relative de structură îşi găsesc o largă aplicare în cercetarea statistică, atît în prelucrarea statistică pentru obţinerea diferiţilor indicatori derivaţi, cît mai ales în procesul de analiză a structurii colectivităţii pe tipuri calitative şi a deplasărilor de structură intervenite de la o perioadă la alta. Dacă au aceeaşi bază de raportare se poate opera cu ele ca şi cînd ar fi mărimi absolute, adică se poate restrînge numărul de grupe, însumînd greutăţile specifice corespunzătoare acestora, sau se pot face cu ele operaţii de scădere. În cazul cînd au baze diferite, greutăţile specifice corespunzătoare acelorași grupe se compară tot sub formă de raport sau se spune că ponderea unei grupe s-a modificat de la $x\%$ la $y\%$.

Şi mai sugestive apar mărimile relative de structură dacă se reprezintă grafic, folosind pătratul, cercul, semicercul sau dreptunghiul.

În cazul pătratului de structură se consideră că suprafaţa este egală cu 100%, deci se construieşte o reţea care cuprinde 100 de pătrăţele egale. Pentru fiecare grupă se va haşura sau colora diferit un număr egal de pătrăţele cu greutatea lor specifică.

Pentru exemplificare se vor reprezenta grafic datele din tabelul 5.2.

Tabelul 5.2

Structura pe medii a populaţiei R.S. România la recensămintele din anii : 1956; 1966; 1977 (% faţă de total)

Populaţia	1956	1966	1977
A	1	2	3
Populaţia urbană	31,3	38,2	47,5
Populaţia rurală	68,7	61,8	52,5
Total	100,0	100,0	100,0

Sursa : Anuarul Statistic al R.S. România, 1977, D.C.S., p. 45.

Scanned with OKEN Scanner

Dacă în tabel ar fi fost date exprimate în mărimi absolute, atunci ele trebuiau transformate în mărimi relative de structură și apoi reprezentate grafic.

Datele din tabelul 5.2 se vor reprezenta în trei pătrate cu laturi egale, în care se vor haşura două grupe (vezi fig. 5.1).

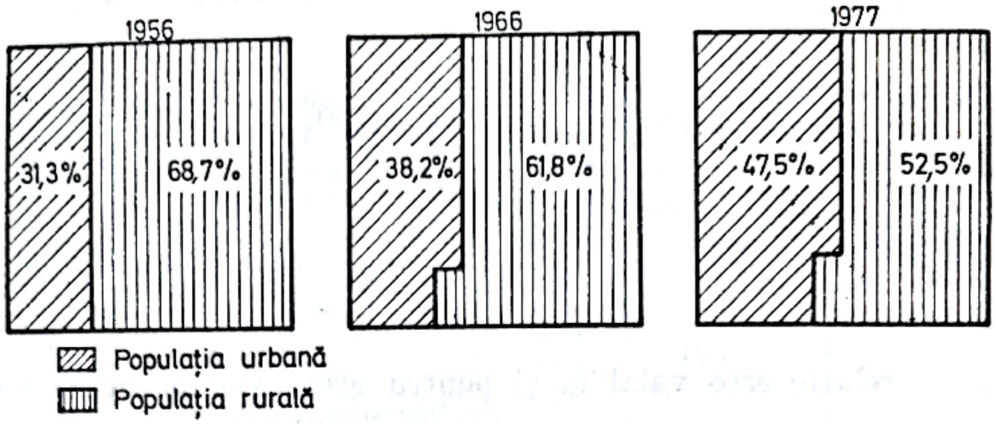


Fig. 5.1. Structura pe medii a populației în R.S. România în anii 1956, 1966 și 1977.

Folosind cercul de structură, se consideră că 360° corespund cu 100%. Considerînd egalitatea 360° = 100%, se calculează apoi cîte grade corespund fiecărei greutăți specifice. Practic, fiecare greutate specifică exprimată în procente se înmulțește cu 3,60° (tabelul 5.3).

Tabelul 5.3

Calculele necesare reprezentării prin cerc a structurii

Anii	Nr. de grade în raport cu structura populației	
	Urbană	Rurală
1956	31,3 × 3,60° = 113°	68,7 × 3,60° = 247°
1966	38,2 × 3,60° = 138°	61,8 × 3,60° = 222°
1977	47,5 × 3,60° = 171°	52,5 × 3,60° = 189°

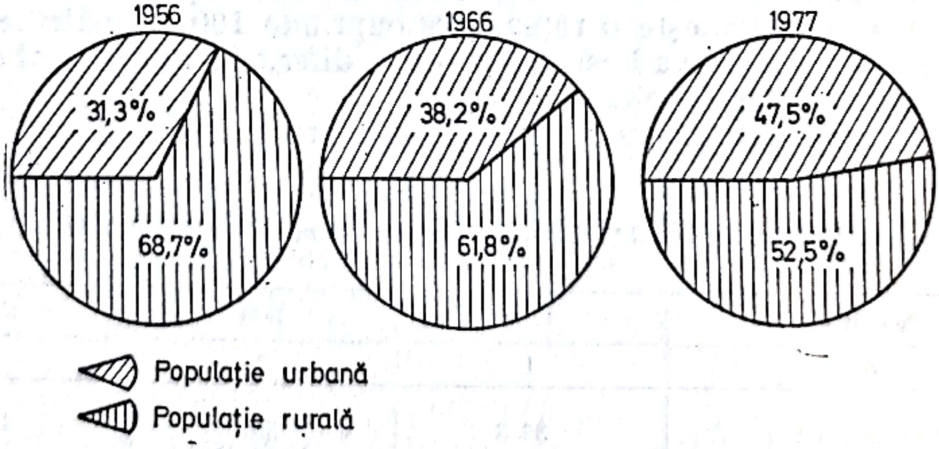


Fig. 5.2. Structura pe medii a populației în R.S. România în anii 1956, 1966 și 1977.

Pentru reprezentare, numărul de grade s-a rotunjit astfel încît să se obțină numere întregi (la cele sub 0,50 se renunță la zecimale, iar la cele peste 0,50 se adaugă o unitate la numărul întreg).

Scanned with OKEN Scanner

O variantă a acestui grafic este cea în care se folosesc *semicercurile*. Prin analogie cu cercul de structură, aici se foloseşte relaţia 180° = 100% şi, de regulă, semicercurile se pun faţă în faţă, la o distanţă foarte mică unul de altul. De exemplu, s-ar putea folosi un astfel de grafic dacă s-ar reprezenta aceeaşi indicatori dar înregistraţi pentru două judeţe.

Diagrama de structură folosind dreptunghiul presupune că cele 100% sînt proporţionale cu înălţimea dreptunghiului. În exemplul considerat se vor construi trei dreptunghiuri egale între ele ca suprafaţă şi se vor separa apoi în cîte două părţi proporţionale cu greutăţile specifice ale celor două grupe de populaţie (vezi fig. 5.3).

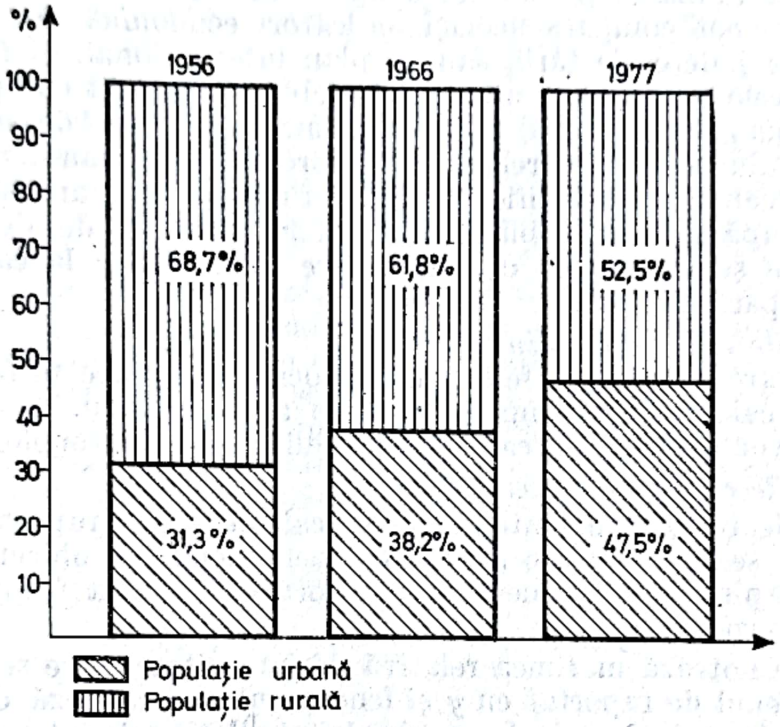


Fig. 5.3. Structura pe medii a populației în R.S. România în anii 1956, 1966 și 1977.

Acest grafic se folosește în special atunci cînd se compară vizual structura mai multor colectivităţi înregistrate în dinamică. Graficul se poate construi atunci în cadranul I din sistemul axelor rectangulare, iar scara de reprezentare, exprimată în procente, se va trece pe axa ordonatelor, avînd aliura unui grafic prin coloane.

Mărimile relative de coordonare

În procesul de cercetare statistică interesează uneori raporturile numerice care există între grupele aceleiaşi colectivitate sau între două sau mai multe colectivităţi de acelaşi fel existente în aceeaşi unitate de timp şi în unităţi teritoriale diferite.

Rezultatul acestor comparări este o mărime relativă de coordonare. Aceste mărimi se exprimă fie sub formă de coeficienţi, fie sub formă de procente.

Scanned with OKEN Scanner

Dacă se notează cu x_A nivelul grupei (colectivității) care se compară și cu x_B nivelul grupei (colectivității) aleasă ca bază de comparare, atunci formula de calcul a mărimilor relative de coordonare este:

$$K_{A/B} = \frac{x_A}{x_B} \cdot 100. \qquad (5.9)$$

Astfel, se pot compara: nivelul producției industriale a grupei A cu nivelul producției industriale a grupei B, numărul populației urbane cu numărul populației rurale, numărul populației ocupate în sfera producției materiale cu numărul populației ocupate în sfera nematerială etc. De asemenea, se pot compara aceiași indicatori economici sau demografici din diferitele județe ale țării, sau pe plan internațional. Și în acest sens problemele cele mai dificile apar cînd trebuie asigurată comparabilitatea datelor pe plan internațional și cînd trebuie aleasă baza de referință teritorială. Calculul mărimilor relative de coordonare pe plan teritorial nu se poate face decît după o verificare critică a datelor care urmează să se raporteze și după ce s-au stabilit unele criterii obiective de alegere a bazei de raportare și a gradului de cuprindere a unităților la care se poate asigura comparabilitatea.

Mărimile relative ale dinamicii

Compararea în timp a fenomenelor social-economice se face, în principal, prin calcularea mărimilor relative ale dinamicii. Aceste mărimi relative se vor prezenta în capitolul „Analiza seriilor cronologice".

Mărimile relative de intensitate

Mărimile relative de intensitate se calculează ca raport între două fenomene ce se găsesc într-o relație de interdependență obiectivă și arată gradul de răspîndire a fenomenului în raport de mărimea fenomenului ales ca bază de raportare.

Dacă se notează mărimea relativă de intensitate, care se calculează, cu x, fenomenul de raportat cu y și fenomenul ales ca bază de raportare cu z, atunci formula de calcul a unei mărimi relative de intensitate va fi:

$$x = \frac{y}{z}. \qquad (5.10)$$

Rezultă clar că, în acest caz, mărimea relativă de intensitate x se exprimă în unități concrete de măsură ale indicatorului y, care revin pe o unitate a indicatorului z.

Mărimile relative de intensitate se pot calcula pe fiecare unitate în parte, pe o grupă de unități sau pe o colectivitate. În cazul în care mărimea relativă de intensitate se calculează din valorile individuale sau parțiale ale unor caracteristici statistice, atunci rezultatele raporturilor pot fi considerate și ele ca valori ale unei variabile statistice aleatoare, care poate fi supusă în continuare unor prelucrări succesive în vederea obținerii unor indicatori medii, de variație, de corelație etc. Pentru a le deosebi de caracteristicile statistice incluse în programul observării, aceste mărimi relative de intensitate se mai numesc și caracteristici derivate, deoarece ele se ob-

Scanned with OKEN Scanner

ţin printr-un calcul statistic, iar centralizarea lor pe grupe şi pe total se face după acelaşi procedeu după care s-au calculat şi nu prin însumare ca în cazul indicatorilor absoluţi.

Mărimile relative de intensitate au o largă aplicabilitate în statistica social-economică. Printre acestea se pot aminti numai cîteva exemple din domeniul statisticii demografice, din domeniul economic şi din domeniul social-cultural.

Din statistica demografică este suficient să se amintească cîţiva coeficienţi demografici generali : coeficientul de mortalitate infantilă; coeficientul de natalitate generală; coeficientul de mortalitate generală etc.

În domeniul economic, mărimile relative de intensitate sînt foarte numeroase, printre ele enumerîndu-se : nivelul productivităţii muncii; retribuţia pe un muncitor; randamentul mediu pe o maşină; recolta la ha, costurile la ha etc.

În domeniul social-cultural se calculează o serie de indicatori statistici ca raport între indicatorii care caracterizează nivelul de trai material sau cultural (consumul total al populaţiei, număr de studenţi, număr de cadre de specialitate, număr de paturi de spital etc.) şi numărul mediu al populaţiei. Aceşti indicatori se exprimă fie în promile ($^0/_{00}$), fie în prodecimile ($^0/_{000}$).

În concluzie, mărimile relative au o utilizare foarte largă în cercetarea statistică. Ele servesc la compararea datelor statistice în scopul măsurării raporturilor cantitative dintre diferitele caracteristici statistice, între fenomene de acelaşi fel sau care se găsesc într-un raport de interdependenţă.

Utilitatea lor este şi mai mare dacă alături de ele sînt prezentate şi mărimile absolute din compararea cărora au rezultat. Folosirea mărimilor relative, în combinaţie cu mărimile absolute din care s-au calculat, este obligatorie atunci cînd ele se referă la baze diferite. Aceasta permite formularea unor concluzii juste cu privire la fenomenele şi procesele social-economice cercetate.

Scanned with OKEN Scanner

MĂRIMILE MEDII, INDICATORII VARIAȚIEI ȘI ASIMETRIEI

Analiza statistică a trăsăturilor esențiale ale fenomenelor de masă, stabilirea tendințelor și regularităților ce apar în producerea lor necesită calcularea anumitor valori tipice, care să fie reprezentative pentru întreaga colectivitate studiată. Determinarea valorilor tipice în statistică se realizează, în principal, cu ajutorul mărimilor medii.

6.1. MĂRIMILE MEDII ÎN STATISTICA SOCIAL-ECONOMICĂ. CONDIȚII DE APLICARE

Mărimile medii sînt utilizate ca instrumente principale de cunoaștere a fenomenelor de masă, deoarece numai pe baza lor se poate exprima ceea ce este comun și general în forma de manifestare a acestor fenomene, în fiecare etapă dată.

De aceea, utilizarea mărimilor medii pentru caracterizarea statistică a fenomenelor sociale și economice de masă implică rezolvarea unor probleme legate de conținutul acestora.

Mărimile medii, făcînd parte din cadrul indicatorilor derivați, se prezintă ca mărimi cu caracter abstract, chiar dacă ele se exprimă în unități concrete de măsură. De exemplu, retribuția medie lunară într-o întreprindere a fost de 2 520,75 lei. Mărimea sa exactă poate să nu fi fost încasată de nici un om din întreprindere, dar ea se încadrează între limitele retribuțiilor individuale din care s-a calculat media. Conținutul acestei valori este însă diferit de la o grupă de personal la alta : pentru muncitorii cu calificare medie această valoare este mai reprezentativă decît pentru cei cu retribuții minime sau maxime din aceeași colectivitate. În consecință, pentru a asigura un conținut cît mai real mediilor calculate, este necesar ca valorile individuale din care se calculează să fie cît mai apropiate ca mărime între ele pentru a putea permite sintetizarea lor într-o singură valoare reprezentativă. Mai mult, ținînd seama de rolul mărimilor medii în cunoașterea legilor care acționează sub formă de tendință — legi specifice fenomenelor de tip colectiv — reiese și mai evident necesitatea de a utiliza la calculul mediilor numai mărimi individuale de același tip calitativ, pentru ca media calculată să poată exprima în mod veridic esența social-economică a acestora.

Un model bazat pe un sistem închegat de categorii economice stabilite sub formă de medii este analiza făcută de Karl Marx societății capitaliste.

Analizînd societatea capitalistă la nivelul întregii economii naționale, Marx și-a bazat concluziile pe interpretarea științifică a unor categorii economice, ce se formează sub formă de mărimi medii, printre care se pot aminti : profitul mediu, rata medie a profitului, rata medie a plusvalorii, durata medie a zilei de lucru și altele. „Munca materializată în va-

Scanned with OKEN Scanner

loare — scria Marx — este muncă de calitate socială medie, deci manifestarea unei forţe de muncă medie"[1]. În acelaşi timp el atrăgea atenţia asupra modului de interpretare a conţinutului acestor valori medii : „...o mărime medie nu există niciodată decît ca medie a unui mare număr de mărimi individuale diferite din aceeaşi specie"[2].

Este evident faptul că el condiţiona folosirea mărimilor medii de existenţa unor fenomene asemănătoare care se prezintă sub forme individuale diferite şi foarte numeroase (deci fenomene de masă), care formează obiectul de cercetare al statisticii. Interpretînd însă aceste categorii cu caracter de medie el arată, în continuare, că în procesul concret al producţiei, fiecare muncitor se poate abate mai mult sau mai puţin de la muncitorul mediu, dar pe ansamblu aceste diferenţe, care pot fi considerate ca erori se compensează reciproc, dînd naştere unor mărimi tipice către care tind să se concentreze cele mai multe valori individuale. Aceasta corespunde, de fapt, interpretării statistice a legii numerelor mari, care presupune că într-un număr suficient de mare de cazuri individuale diferite se manifestă o anumită tendinţă de distribuţie a erorilor[3], la care apariţia unor valori întîmplătoare nu poate modifica tendinţa lor de concentrare către anumite valori, faţă de care s-a asigurat o compensare reciprocă a abaterilor produse într-un sens sau altul.

Principiul omogenităţii datelor individuale din care se poate calcula o mărime medie este frecvent amintit şi în operele lui V.I. Lenin, în legătură cu modul de interpretare a categoriilor economice ce apar ca valori medii. El sublinia faptul că prin calcule statistice se pot stabili fie medii cu caracter reprezentativ, fie medii cu caracter fictiv. Atunci cînd media este calculată pentru o colectivitate cu o structură eterogenă, ea are un caracter de medie globală, lipsită de conţinut şi care, în loc să dezvăluie esenţa fenomenului, duce la estomparea diferenţierilor calitative ce apar în mod obiectiv în cadrul acestor colectivităţi. Analizînd datele din diferitele publicaţii din vremea sa, din mai multe ţări capitaliste, Lenin arăta caracterul fictiv a unor medii globale, calculate pe întreaga populaţie (venitul mediu pe o persoană etc.).

Pornind de la aceste considerente, este necesar ca la aplicarea metodei mediilor să se ţină seama de gradul de omogenitate al colectivităţii cercetate. În cazul în care colectivitatea este eterogenă se vor calcula numai medii parţiale, iar în măsura în care media pe întregul ansamblu apare ca o sinteză a mediilor parţiale are sens economic şi calculul mediei generale. Reiese de aici că metoda mediilor trebuie utilizată împreună cu metoda grupărilor prin aplicarea căreia unităţile de la care s-au cules datele se separă în grupe omogene după variaţia uneia sau a mai multor caracteristici esenţiale. Rezultă că la aplicarea metodei mediilor este necesar să se respecte următoarele condiţii :

Calculul mediilor să se bazeze pe folosirea unui număr mare de cazuri individuale diferite sub care s-a înregistrat caracteristica, a căror variaţie este întîmplătoare în raport cu întreaga masă de fenomene studiate.

---

[1] K. Marx, *Capitalul*, vol. I, Ed. P.M.R., 1948, p. 305.
[2] Idem.
[3] În sens statistic, aceste erori se interpretează ca abateri între valorile individuale şi valorile tipice care se formează în mod obiectiv pe baza unei legi de distribuţie.

Scanned with OKEN Scanner

Cuprinderea în calculul mediei, exhaustiv sau reprezentativ, a cazurilor individuale care formează colectivitatea sau grupele aparţinînd aceleiaşi esenţe social-economice, pentru a putea asigura includerea tuturor factorilor esenţiali şi întîmplători care determină variaţia caracteristicilor şi care să permită un cîmp larg de acţiune legii numerelor mari, în vederea asigurării compensării într-un sens sau altul a abaterilor individuale de la tendinţa lor centrală. Atenţie deosebită trebuie acordată, în special, calculării mărimilor medii pe baza datelor obţinute din observări parţiale. Acestea se pot extinde pentru întreaga colectivitate numai în măsura în care se îndeplineşte condiţia de reprezentativitate, adică observările parţiale sînt organizate prin sondaj selectiv. În acest caz este necesar să se verifice prin procedee statistice măsura în care mărimea medie de selecţie prezintă stabilitate şi este reprezentativă pentru întreaga masă de fenomene[1].

Asigurarea omogenităţii valorilor individuale pentru care se calculează media, prin analiza gradului şi formei de distribuţie a abaterilor dintre valorile individuale şi valorile tipice, realizată prin interpretarea vizuală a seriei de distribuţie şi a reprezentării sale grafice.

Alegerea formei de medie care corespunde cel mai bine formei de variaţie a caracteristicii studiate, pentru ca mărimea medie să aibă un conţinut obiectiv, măsurînd în mod veridic fenomenele social-economice cercetate.

6.2. FELURILE MEDIILOR UTILIZATE ÎN STATISTICA SOCIAL-ECONOMICĂ

În practica statistică, pentru determinarea valorii tendinţei centrale a unei serii statistice se pot folosi mai multe feluri de medii, printre care cele mai frecvente se întîlnesc : media aritmetică ; media armonică ; media pătratică ; media geometrică ; media cronologică[2].

Toate aceste feluri de medii se pot calcula ca medii simple sau ca medii ponderate. Mediile simple se folosesc atunci cînd numărul variantelor sub care s-a înregistrat caracteristica este egal cu numărul unităţilor la care s-a făcut observarea, sau atunci cînd frecvenţele tuturor variantelor sînt egale între ele şi, deci, se pot simplifica. Mediile ponderate se folosesc atunci cînd aceeaşi variantă se întîlneşte la mai multe unităţi de observare, deci, fiecărei variante a caracteristicii i se poate ataşa o frecvenţă.

În statistica social-economică, cel mai frecvent se foloseşte media aritmetică. Din această cauză, în mod curent, cînd se vorbeşte despre o mărime medie, se înţelege că s-a calculat o medie aritmetică. În cazul în care s-a calculat un alt fel de medie este necesar să se specifice felul ei (armonică, pătratică etc.).

---

[1] Această problemă va fi tratată pe larg în capitolul „Cercetarea selectivă".

[2] Din considerente didactice media cronologică se va prezenta la capitolul „Analiza seriilor cronologice".

Scanned with OKEN Scanner

6.2.1. Media aritmetică simplă și ponderată

Media aritmetică este rezultatul sintetizării într-o singură expresie numerică a tuturor nivelurilor individuale observate, obținută prin raportarea valorii totalizate a caracteristicii la numărul total al unităților.

În sens statistic, media aritmetică calculată pentru o colectivitate statistică este valoarea care s-ar fi înregistrat dacă toți factorii ar fi influențat în mod constant în toate cazurile înregistrate.

Dacă se notează cu x variabila pentru care se calculează media, atunci ea poate lua valori de la $x_1 \ldots x_n$. În mod corespunzător se va nota mărimea cu $\bar{x}$.[1] Dacă media aritmetică este o valoare tipică pentru un ansamblu de fenomene înseamnă că ea reprezintă acea valoare cu care, dacă s-ar înlocui toți termenii individuali ai seriei și s-ar face suma lor, această valoare ar fi egală cu suma termenilor reali din care s-a calculat media, adică:

$$x_1 + x_2 + x_3 + \ldots + x_n = \sum_{i=1}^{n} x_i \qquad (6.1)$$

$$\bar{x} + \bar{x} + \bar{x} + \ldots + \bar{x} = \sum_{i=1}^{n} x_i, \qquad (6.2)$$

n reprezintă numărul termenilor pentru care se calculează media: $i = 1, 2, 3, \ldots, n$.

În relația (6.2) $\bar{x}$ fiind o mărime constantă, suma acestei valori este egală cu produsul dintre constantă și numărul termenilor pentru care se însumează, adică:

$$n\bar{x} = \sum_{i=1}^{n} x_i \quad (6.3), \text{ de unde } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (6.4)$$

Deci, media aritmetică simplă se calculează ca raport între suma nivelurilor individuale sub care s-a înregistrat caracteristica și numărul cazurilor individuale luate în observare. Dacă nu se dispune de date individuale, ci numai de datele centralizate ale caracteristicii, media se poate calcula direct împărțind valoarea totalizată a caracteristicii la numărul total al unităților la care s-a făcut centralizarea.

De exemplu, dacă producția unei echipe de muncitori într-o zi a fost de 3 000 bucăți și în echipă sînt 10 muncitori, înseamnă că producția medie

[1] Pentru valoarea mediei se mai folosește și notația $M_{(x)}$.

Scanned with OKEN Scanner

zilnică ce revine pe un muncitor din echipa respectivă este de :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3\,000}{10} = 300 \text{ buc/muncitor.}$$

Deci, în medie, în ziua respectivă se consideră că fiecare muncitor din această echipă a realizat cîte 300 bucăţi. Media se exprimă în aceleaşi unităţi de măsură ca şi nivelurile individuale ale caracteristicii din care s-a calculat.

Într-o colectivitate statistică se întîlnesc foarte rar cazuri în care numărul variantelor coincide cu numărul unităţilor. De regulă, fenomenele de masă sînt numeroase şi aceeaşi valoare individuală poate fi întîlnită de mai multe ori. În acest caz, pentru a putea cuprinde în calcul toate valorile individuale, trebuie să se ţină seama şi de frecvenţa lor de apariţie, iar seria pentru care trebuie să se calculeze se prezintă astfel :

$$x_1,\ x_2 \ldots x_m$$
$$f_1,\ f_2 \ldots f_m$$

în care :

$f_{1,2}\ldots m$ reprezintă frecvenţa sau ponderea fiecărei variante ;

m — numărul variantelor şi, respectiv, ponderilor pentru o serie de frecvenţe. Dacă n este numărul unităţilor colectivităţii statistice, în cazul seriei de distribuţie de frecvenţe $m < n$.

În acest caz, pentru a calcula media, valoarea totalizată a caracteristicii se obţine pe baza unor valori totalizate parţial sub forma unor produse de frecvenţe (fiecare variantă a seriei se înmulţeşte cu frecvenţa sa), adică :

$$x_1 f_1 + x_2 f_2 + \ldots + x_m f_m = \sum_{i=1}^{m} x_i f_i \qquad (6.5)$$

Folosind aceeaşi proprietate ca şi în cazul precedent, se obţine :

$$\bar{x} f_1 + \bar{x} f_2 + \ldots + \bar{x} f_m = \sum_{i=1}^{m} x_i f_i. \qquad (6.6)$$

Dînd factor comun pe $\bar{x}$ în relaţia (6.6) se obţine :

$$\bar{x}(f_1 + f_2 + \ldots + f_m) = \sum_{i=1}^{m} x_i f_i \qquad (6.7)$$

adică :

$$\bar{x} \sum_{i=1}^{m} f_i = \sum_{i=1}^{m} x_i f_i \qquad (6.8)$$

Scanned with OKEN Scanner

de unde:

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i} \tag{6.9}$$

Această formulă a mediei aritmetice ponderate se aplică pentru o serie de distribuţie de frecvenţă şi se bazează pe însumarea produselor de frecvenţă $\left(\sum_{i=1}^{m} x_i f_i\right)$ şi raportarea rezultatului obţinut la totalul frecvenţelor $\left(\sum_{i=1}^{m} f_i\right)$.

6.2.1.1. PROPRIETĂŢILE MEDIEI ARITMETICE

Media aritmetică posedă o serie de proprietăţi dintre care unele au utilitate practică pentru calculul şi interpretarea valorii ei. De aceea nu vor fi prezentate aici decît cele care se utilizează la aplicarea mediilor în statistica social-economică.

1° Într-un şir de valori egale media acestora este egală cu fiecare dintre ele, indiferent dacă este o serie simplă sau o serie de frecvenţe, adică:

$$x_1 = x_2 = x_3 = \ldots = x_n = x_c,$$

în care:

x_c reprezintă valoarea constantă a caracteristicii.

Deci:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_c}{n} = \frac{nx_c}{n} = x_c \tag{6.10}$$

Pentru o serie de distribuţie de frecvenţe media va fi:

$$\bar{x} = \frac{\sum_{i=1}^{m} x_c f_i}{\sum_{i=1}^{m} f_i} = \frac{x_c \sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} f_i} = x_c \tag{6.11}$$

2° Mărimea mediei aritmetice este totdeauna o valoare cuprinsă în intervalul de variaţie al caracteristicii, adică:

$$x_{\min} < \bar{x} < x_{\max} \tag{6.12}$$

Scanned with OKEN Scanner

Fie $x_{max} = \max(x_1, x_2, \ldots, x_n)$ și $x_{min} = \min(x_1, x_2, \ldots, x_n)$, avem :

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} < \frac{x_{max} \Sigma f_i}{\Sigma f_i} = x_{max}; \quad \text{deci } \bar{x} < x_{max}$$

Analog :

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} > \frac{x_{min} \Sigma f_i}{\Sigma f_i} = x_{min}; \quad \text{deci } \bar{x} > x_{min}$$

și în final :

$$x_{min} < \bar{x} < x_{max}$$

3° În cazul unei serii de distribuție de frecvențe media se încadrează între valorile extreme ale variabilei, oscilînd în jurul termenului cu frecvența cea mai mare.

Aceste proprietăți (1°; 2°, 3°) servesc pentru o primă verificare a exactității mediei calculate. Dacă valoarea medie a unei serii statistice se plasează în afara limitelor extreme ale variantelor seriei și prezintă o abatere foarte mare față de termenul cu frecvența cea mai mare, atunci este un prim indiciu că s-a strecurat o eroare de calcul și, deci, media trebuie recalculată.

4° Într-o serie statistică, suma algebrică a tuturor abaterilor individuale ale termenilor seriei de la media lor aritmetică este egală cu zero.

În cazul unei serii simple, această proprietate se exprimă astfel :

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0. \tag{6.13}$$

Într-adevăr :

$$\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}_i = \sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - \frac{n \sum_{i=1}^{n} x_i}{n} = 0$$

În cazul seriei de frecvențe, compensarea reciprocă a abaterilor trebuie să se facă pentru toate unitățile respective :

$$\sum_{i=1}^{m} (x_i - \bar{x}) f_i = 0 \tag{6.14}$$

adică :

$$\sum_{i=1}^{m} (x_i - \bar{x}) f_i = \sum_{i=1}^{m} x_i f_i - \sum_{i=1}^{m} \bar{x} f_i = \sum_{i=1}^{m} x_i f_i - \bar{x} \sum_{i=1}^{m} f_i =$$

$$= \sum_{i=1}^{m} x_i f_i - \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i} \cdot \sum_{i=1}^{m} f_i = \sum_{i=1}^{m} x_i f_i - \sum_{i=1}^{m} x_i f_i = 0.$$

Scanned with OKEN Scanner

5° Într-o serie statistică, dacă se micşorează sau se măresc toţi termenii seriei cu acecaşi constantă (a) şi se face media aritmetică a noilor termeni, această medie este mai mică sau mai mare decît media seriei iniţiale cu constanta a, adică:

— seria iniţială notată cu (x_i) va fi:

$$x_1,\ x_2,\ \ldots,x_n$$

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n};$$

— seria a cărei termeni au fost măriţi sau micşoraţi cu o constantă a, notată cu (x_i'), va fi:

$$x_1' = x_1 \pm a;\quad x_2' = x_2 \pm a;\quad \ldots\ x_n' = x_n \pm a$$

$$\bar{x} = \frac{(x_1 \pm a) + (x_2 \pm a) + \ldots + (x_n \pm a)}{n} = \frac{\sum_{i=1}^{n}(x_i \pm a)}{n} =$$

$$= \frac{\sum_{i=1}^{n} x_i \pm \sum_{i=1}^{n} a}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \pm \frac{na}{n} = \bar{x} \pm a.$$

Deci:

$$\bar{x}' = \bar{x} \pm a. \tag{6.15}$$

6° Într-o serie statistică, dacă se amplifică sau se simplifică toţi termenii seriei cu un factor constant k şi se face media noilor termeni, media astfel obţinută va fi de k ori mai mică decît media seriei iniţiale:

— seria iniţială notată cu (x_i) va fi:

$$x_1;\ x_2;\ x_3;\ \ldots;\ x_n;\ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n};$$

— seria simplificată[1] de k ori, notată de pildă cu (x_i''), deci:

$$x_1'' = \frac{x_1}{k};\quad x_2'' = \frac{x_2}{k};\quad \ldots\ x_n'' = \frac{x_n}{k};$$

$$\bar{x}'' = \frac{\frac{x_1}{k} + \frac{x_2}{k} + \ldots + \frac{x_n}{k}}{n} = \frac{\Sigma \frac{x_i}{k}}{n}.$$

[1] Pentru verificare s-a folosit numai operaţia de simplificare, în mod analog se poate demonstra această proprietate şi pentru amplificarea termenilor seriei iniţiale.

Scanned with OKEN Scanner

Dînd factor comun pe $\frac{1}{k}$

$$\bar{x}'' = \frac{\frac{1}{k}\sum_{i=1}^{n} x_i}{n} = \frac{1}{k}\bar{x}, \text{ deci } \bar{x}'' = \frac{\bar{x}}{k} \tag{6.16}$$

Proprietăţile de la punctele 5° şi 6° servesc în statistică pentru aplicarea unor procedee de calcul simplificat; ele se pot folosi separat sau împreună.

7° Dacă într-o serie de distribuţie se reduc proporţional toate frecvenţele, media calculată pe baza noilor frecvenţe rămîne neschimbată, respectiv :

— media seriei iniţiale :

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_m f_m}{f_1 + f_2 + \ldots + f_m} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i};$$

— media seriei cu frecvenţe reduse, printr-un factor constant (c);

$$\bar{x} = \frac{x_1 \frac{f_1}{c} + x_2 \frac{f_2}{c} + \ldots + x_m \frac{f_m}{c}}{\frac{f_1}{c} + \frac{f_2}{c} + \ldots + \frac{f_m}{c}} = \frac{\sum_{i=1}^{m} x_i \frac{f_i}{c}}{\sum_{i=1}^{m} \frac{f_i}{c}}. \tag{6.17}$$

Dînd factor comun pe $\frac{1}{c}$ atît la numărător, cît şi la numitorul fracţiei :

$$\frac{\frac{1}{c}\sum_{i=1}^{m} x_i f_i}{\frac{1}{c}\sum_{i=1}^{m} f_i} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i} = \bar{x}. \tag{6.18}$$

Această proprietate se aplică ca mijloc de simplificare a calculului mediei atunci cînd frecvenţele admit acelaşi coeficient de simplificare. Cel mai frecvent însă, această proprietate este folosită la calculul mediei pe baza frecvenţelor relative (a se vedea paragraful 6.2.1.2).

8°. Într-o serie de variaţie cu frecvenţe egale, media aritmetică ponderată se transformă în medie aritmetică simplă.

Scanned with OKEN Scanner

Dacă se notează cu f_c — frecvența constantă, atunci există relația:

$$f_1 = f_2 = f_3 = \ldots = f_n = f_c$$

$$\bar{x} = \frac{x_1 f_c + x_2 f_c + \ldots + x_m f_c}{f_c + f_c + \ldots + f_c} = \frac{\sum_{i=1}^{m} x_i f_c}{\sum_{i=1}^{m} f_c}. \qquad (6.19)$$

Dînd factor comun pe f_c la numărător, se obține:

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i f_c}{\sum_{i=1}^{m} f_c} = \frac{f_c' \sum_{i=1}^{m} x_i}{m f_c'} = \frac{\sum_{i=1}^{m} x_i}{m}. \qquad (6.20)$$

9° Media generală calculată pe baza distribuției de frecvențe a colectivității generale este egală cu media mediilor parțiale calculate pe baza seriilor de distribuție componente.

Aceasta înseamnă că seriile parțiale nu admit o teoremă de adițiune, respectiv media generală se formează ca sinteză a mediilor parțiale și este cuprinsă între cea mai mică și cea mai mare medie parțială.

Media generală nefiind o sumă a mediilor parțiale, ci o sinteză a acestora, înseamnă că pe întreaga colectivitate are loc un proces de compensare a abaterilor dintre mediile parțiale și cea generală. (Formulele de calcul și relațiile dintre ele se pot urmări în paragraful în care se tratează regula adunării dispersiilor).

10° Media aritmetică a sumei a două variabile întîmplătoare x, y este egală cu suma mediilor celor două variabile luate în calcul:

$$\overline{x + y} = \bar{x} + \bar{y}, \qquad (6.21)$$

deoarece:

$$\frac{(x_1 + y_1) + (x_2 + y_2) + \ldots + (x_n + y_n)}{n} = \frac{\sum_{i=1}^{n} (x_i + y_i)}{n} =$$

$$= \frac{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} x_i}{n} + \frac{\sum_{i=1}^{n} y_i}{n} = \bar{x} + \bar{y}.$$

Scanned with OKEN Scanner

Această proprietate se folosește în statistica economică atunci cînd în calcule apar variabile statistice aleatoare complexe, care se formează ca sumă a două sau mai multe variabile aleatoare independente. Astfel, dacă se consideră cheltuielile de producție ca o variabilă aleatoare complexă (z) formată din suma a două variabile aleatoare independente: cheltuielile fixe de producție (x) și cheltuielile variabile de producție (y), atunci cheltuielile medii pe total colectivitate sînt egale cu cheltuielile medii de producție fixe și cheltuielile medii variabile: $\bar{z} = \bar{x} + \bar{y}$.

În general, această proprietate se poate enunța astfel: valoarea medie a sumei mai multor variabile aleatoare independente este egală cu suma valorilor medii ale tuturor variabilelor independente luate în calcul.

11° Media produsului variabilelor întîmplătoare independente x, y este egală cu produsul mediilor celor două variabile x, y.

Pentru aceasta se pornește de la ipoteza că dacă există două variabile aleatoare independente a căror produs poate fi considerat o nouă variabilă aleatoare complexă, atunci media acesteia este egală cu produsul celor două medii calculate independent, adică:

$$\overline{xy} = \bar{x} \cdot \bar{y}. \tag{6.22}$$

Analitic, cele trei variabile (două variabile factoriale și una complexă) se prezintă astfel:

x	x_1	x_2	...	x_n
y	y_1	y_2	...	y_n
xy	x_1y_1	x_1y_2	...	x_1y_n
	x_2y_1	x_2y_2	...	x_2y_n
	⋮	⋮		⋮
	x_ny_1	x_ny_2		x_ny_n

pentru care se pot calcula mediile:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}; \quad \bar{y} = \frac{y_1 + y_2 + \ldots + y_n}{n} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$\overline{xy} = \frac{x_1y_1 + x_1y_2 + \ldots + x_1y_n + x_2y_1 + x_2y_2 + \ldots + x_2y_n +}{nn}$$

$$\frac{+ \ldots + x_ny_1 + x_2y_2 \ldots x_ny_n}{nn} = \frac{x_1(y_1 + y_2 + \ldots + y_n) +}{nn}$$

$$\frac{+ x_2(y_1 + y_2 + \ldots + y_n) + \ldots + x_n(y_1 + y_2 + \ldots + y_n)}{nn}$$

Scanned with OKEN Scanner

$$= \frac{x_1 \sum_{i=1}^{n} y_i + x_2 \sum_{i=1}^{n} y_i + \ldots + x_n \sum_{i=1}^{n} y_i}{nn} = \frac{\sum_{i=1}^{n} y_i (x_1 + x_2 + \ldots + x_n)}{nn} =$$

$$= \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_n}{nn} = \frac{\sum_{i=1}^{n} y_i}{n} \cdot \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \cdot \bar{y}.$$

Această proprietate se foloseşte frecvent în cercetarea statistică, în special în diferite calcule în care se folosesc variabile statistice care se pot descompune în produs de mai mulţi factori, ca de exemplu, volumul producţiei, care este egal cu produsul dintre nivelul de productivitate a muncii şi numărul de muncitori şi altele.

Şi această proprietate poate fi generalizată astfel : media produsului mai multor variabile întîmplătoare, două cîte două, este egală cu produsul mediilor fiecărei variabile.

6.2.1.2. CALCULUL MEDIEI ÎNTR-O SERIE DE DISTRIBUŢIE DE FRECVENŢE PE INTERVALE DE GRUPARE

O serie statistică pe intervale de grupare poate fi prezentată cu frecvenţe absolute sau cu frecvenţe relative. Spre a prezenta un model de calcul al mediei, pentru o serie de intervale, se va folosi exemplul din tabelul 6.1.

Tabelul 6.1

Gruparea muncitorilor unei întreprinderi după retribuţia lunară

Grupe de muncitori după mărimea retribuţiei	Numărul muncitorilor	
	în mărime absolută	în % faţă de total
A	1	2
sub 1 500 lei	48	8,0
1 500—1 800	72	12,0
1 800—2 100	108	18,0
2 100—2 400	123	20,5
2 400—2 700	105	17,5
2 700—3 000	84	14,0
peste 3 000 lei	60	10,0
	Total 600	100,0

Fiind o serie de repartiţie a frecvenţelor trebuie să se folosească formula mediei aritmetice ponderate (6.9).

În cazul prezentat, variabila x este mărimea retribuţiei lunare, iar ponderea este numărul de muncitori corespunzător fiecărui interval şi care este exprimată atît în mărimi absolute, cît şi în mărimi relative.

Scanned with OKEN Scanner

Pentru a calcula valoarea mediei este necesar să se stabilească valoarea retribuției lunare corespunzătoare fiecărui interval.

Considerînd, în mod convențional, că în cadrul fiecărui interval frecvențele s-ar distribui în mod uniform, înseamnă că valoarea medie a caracteristicii în fiecare grupă va fi egală cu media aritmetică simplă a celor două limite. Cu cît intervalele sînt mai mici, cu atît mărimea centrului de interval este mai reprezentativă pentru caracterizarea grupei respective.

Mărimea centrului de interval se mai poate calcula și adăugînd la limita inferioară a intervalului jumătate din mărimea lui. Aceste centre de interval vor fi notate cu $x_1, x_2, \ldots, x_m$.

De exemplu, pentru intervalul al doilea, centrul de interval va fi notat cu x_2 și se poate obține:

$$x_2 = \frac{1\,500 + 1\,800}{2} = 1\,650, \quad \text{sau} \quad x_2 = 1\,500 + \frac{200}{2} = 1\,650\,^{1}.$$

În cazul în care se constată că intervalele sînt egale între ele este suficient să se calculeze mărimea primului centru al intervalului și apoi să se determine o progresie aritmetică crescătoare cu rația egală cu mărimea intervalului de grupare. În cazul intervalelor neegale, calculul fiecărui centru de interval este obligatoriu, deoarece creșterea acestora este neuniformă.

În sfîrșit, dacă seria prezintă și intervale deschise este necesar ca înainte de a trece la stabilirea centrelor de interval să se închidă intervalele care nu au ambele limite precizate. Închiderea intervalelor se face presupunînd că ele vor avea aceeași mărime ca și cele alăturate.

În exemplul prezentat intervalele cu limitele precizate sînt egale, deci se presupune că și cele două intervale extreme s-ar distribui după aceeași mărime (300 lei); deci:

— primul interval va fi: 1 200—1 500 lei;

— ultimul interval va fi: 3 000—3 300 lei.

La calculul mediei se pot folosi frecvențele absolute sau frecvențele relative.

Pentru sistematizarea calculului mediei se vor înscrie datele într-un tabel de calcul, determinînd pe rînd elementele necesare obținerii valorii medii (x, xf); (x, $xf\%$).

Avînd toate elementele de calcul (tab. 6.2) valoarea medie cu frecvențe absolute va fi:

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i} = \frac{1\,367\,100}{600} = 2\,278{,}50 \text{ lei/muncitor.}$$

---

[1] Cînd intervalele sînt cu limite precizate (adică limita superioară nu se repetă ca limită inferioară a intervalului următor), atunci este recomandabil să se folosească cel de-al doilea procedeu.

Scanned with OKEN Scanner

Tabelul 6.2

Calculul mediei aritmetice ponderate pentru o serie de intervale egale

Grupe de retribuţii	Numărul muncitorilor (f_i)	Centrul (mijlocul) intervalului (x_i)	Produse de frecvenţă	
			$(x_i f_i)$	$x_i f_i\%$
0	1	2	3	4
1 200—1 500	48	1 350	64 800	10 800
1 500—1 800	72	1 650	118 800	19 800
1 800—2 100	108	1 950	210 600	35 100
2 100—2 400	123	2 250	276 750	46 125
2 400—2 700	105	2 550	267 750	44 625
2 700—3 000	84	2 850	239 400	39 900
3 000—3 300	60	3 150	189 000	31 500
Total	$\sum_{i=1}^{7} f_i = 600$	—	$\sum_{i=1}^{7} x_i f_i = 1\,367\,100$	$\sum_{i=1}^{7} x_i f_i\% = 227\,850.$

Folosind frecvenţele relative, media va fi:

$$\frac{\sum_{i=1}^{m} x_i f_i\%}{100} = \frac{227\,850}{100} = 2\,278{,}50 \text{ lei/muncitor.}$$

Deci, în cazul seriei de distribuţie de frecvenţe cu intervale de grupare, calculul mediei se face astfel: se stabilesc centrele de interval care se notează cu x_i, apoi se fac produsele de frecvenţă dintre fiecare centru de interval şi frecvenţa sa şi pe baza acestor date se calculează media, raportînd suma produselor de frecvenţă la suma frecvenţelor.

De reţinut că valoarea medie calculată dintr-o serie de distribuţie cu intervale de grupare nu este o medie exactă, deoarece valorile centrale ale intervalelor de grupare nu sînt medii exacte pentru grupele de unităţi constituite după caracteristica de grupare studiată. Din această cauză media calculată din seria de distribuţie cu intervale este o medie aproximativă. Ea se apropie cu atît mai mult de media calculată din datele individuale centralizate, cu cît frecvenţele sînt mai mari, simetric distribuite, iar mărimea intervalelor de grupare este mai mică.

6.2.1.3. CALCULUL SIMPLIFICAT AL MEDIEI ARITMETICE PONDERATE ÎN SERIILE DE INTERVALE

După cum se poate observa, pentru a calcula media într-o serie de intervale trebuie să se facă calcule destul de numeroase. Cu cît frecvenţele sînt mai numeroase, cu atît calculul produselor de frecvenţă necesită mai mult timp. Pentru a reduce din dificultatea calculului se folosesc în practică anumite procedee de calcul simplificat bazate pe unele proprietăţi ale mediilor.

Scanned with OKEN Scanner

Calculul simplificat în seriile cu intervale egale

Pe baza proprietății mediilor (5°) se pot micșora cu o constantă toate centrele de interval și apoi să se calculeze media acestora. Pentru a ajunge la valoarea medie a seriei inițiale trebuie să se facă operația inversă, adică să i se adune constanta respectivă, deci:

$$\bar{x} = \frac{\sum_{i=1}^{m} (x_i - a) f_i}{\sum_{i=1}^{m} f_i} + a. \qquad (6.23)$$

Este avantajos să se aleagă ca valoare pentru a unul dintre centrele de interval; în dreptul acestuia valoarea lui x va fi egală cu 0 și acest centru de interval poate fi considerat ca origine arbitrară pentru variația caracteristicii la care se calculează media (vezi tabelul 6.3, col. 3). Dacă centrul de interval ales ca origine arbitrară este situat în interiorul seriei, atunci, prin aplicarea proprietății amintite, se vor obține pentru (x) valori pozitive (pentru centrele de interval mai mari) și valori negative (pentru centrele de interval mai mici). Însumînd produsele de frecvență pozitive și negative se va reduce și mai mult calculul mediei.

Se observă că dacă intervalele sînt egale (vezi tabelul 6.3), centrele de interval astfel micșorate se pot măsura în unități de mărime a intervalului de grupare. Deci, toate pot fi simplificate cu un coeficient egal cu mărimea intervalului de grupare. Aplicînd proprietatea mediilor (6°), termenii seriei se pot simplifica cu un coeficient K, media se va reduce de K ori; pentru a ajunge la media seriei inițiale, se va înmulți cu K, deci:

$$\bar{x} = \frac{\sum_{i=1}^{m} \frac{x_i}{K} f_i}{\sum_{i=1}^{m} f_i} . K. \qquad (6.24)$$

Este evident că aceste proprietăți se pot folosi separat sau împreună, cum este în cazul seriilor de variație cu intervale egale de grupare. În cazul combinării celor două proprietăți, formula de calcul simplificat al mediei devine:

$$\bar{x} = \frac{\sum_{i=1}^{m} \left(\frac{x_i - a}{K}\right) f_i}{\sum_{i=1}^{m} f_i} . K + a. \qquad (6.25)$$

Scanned with OKEN Scanner

Pentru exemplificare se va utiliza aceeaşi serie de distribuţie de frecvenţe :

$$a = 2\,250;$$

$$K = 300.$$

$$\bar{x} = \frac{\sum_{i=1}^{m}\left(\frac{x_i - a}{k}\right) f_i}{\sum_{i=1}^{m} f_i} \cdot K + a = \frac{+57}{600} \cdot 300 + 2\,250 =$$

$$= 2\,278{,}50 \quad \text{lei/muncitor.}$$

Tabelul 6.3

Calculul simplificat al mediei aritmetice ponderate pentru o serie cu intervale egale

Grupe de retribuţii	Nr. muncitorilor f_i	Centrul intervalului x_i	$x_i - a$	$\frac{x_i - a}{k}$	$\frac{x_i - a}{k} f_i$
0	1	2	3	4	5
1 200—1 500	48	1 350	—600	—3	—144
1 500—1 800	72	1 650	—400	—2	—144
1 800—2 100	108	1 950	—200	—1	—108
2 100—2 400	123	2 250	0	0	0
2 400—2 700	105	2 550	200	1	105
2 700—3 000	84	2 850	400	2	168
3 000—3 300	60	3 150	600	3	180
	600	—	—	—	+ 57
	$\sum_{i=1}^{7} x_i f_i$	—	—	—	$\sum_{i=1}^{7}\left(\frac{x_i - a}{k}\right) f_i$

Pe scurt, calculul simplificat al mediei presupune rezolvarea următoarelor operaţii :

— Verificarea mărimii tuturor intervalelor de grupare. În cazul cînd intervalele de grupare sînt egale între ele, atunci se poate aplica calculul simplificat obţinut pe baza combinării ambelor proprietăţi (micşorarea cu a şi împărţirea cu K);

— Calcularea centrelor de interval notate cu x. În cazul cînd seria este cu intervale deschise, se închid intervalele respective, după mărimea intervalelor alăturate;

— Alegerea unui centru de interval care să fie egal cu a. Pe baza proprietăţilor mediilor, într-o serie de repartiţie de frecvenţe valoarea mediei oscilează în jurul termenului cu frecvenţa cea mai mare. De aceea este avantajos să se aleagă ca valoare pentru a, valoarea centrului de interval cu frecvenţa maximă şi care, de regulă, se găseşte în centrul seriei. Dacă seria este asimetrică, atunci se poate alege un a egal cu al centrului de interval, plasat în mijlocul seriei, chiar dacă nu are frecvenţa cea mai mare;

Scanned with OKEN Scanner

— Alegerea coeficientului K, care, în cazul intervalelor egale, este egal cu mărimea acestora. Dacă intervalele sînt egale şi K = mărimea intervalului, atunci este suficient să se calculeze doar valoarea centrului de interval pe care se ia egal cu a; în dreptul lui se înscrie zero şi se calculează direct coloana $\left(\frac{x-a}{k}\right)$ astfel: de la originea arbitrară se trec, pentru intervalele mai mici decît valoarea aleasă pentru a, numerele negative în ordinea descreşterii lor: -1; -2; -3; etc., iar pentru intervalele mai mari numerele întregi pozitive în ordinea crescătoare: $+1$; $+2$; $+3$; etc., pînă se acoperă toate intervalele de ambele părţi ale axei numerelor întregi;

— Efectuarea produselor de frecvenţă $\left(\frac{x-a}{k}\right)$ pe fiecare interval în parte şi însumarea lor algebrică. Dacă $a < \bar{x}$, atunci produsele de frecvenţă cu semn negativ vor predomina şi, deci, media noilor termeni ai seriei va fi mai mică decît media seriei iniţiale; dacă $a > \bar{x}$, atunci din însumarea algebrică a produselor de frecvenţă se va obţine o valoare pozitivă care se adaugă la valoarea lui a şi se va obţine media. Dacă $a = \bar{x}$, atunci produsele de frecvenţă se vor compensa reciproc şi toată fracţia $\frac{\sum_{i=1}^{m}\left(\frac{x_i-a}{k}\right)f_i}{\sum_{i=1}^{m} f_i} \cdot K$ este egală cu zero.

— Aflarea mediei prin refacerea operaţiilor inverse, efectuate în ordine inversă, respectiv înmulţirea cu coeficientul K şi adunarea cu valoarea lui a;

— Dacă media se determină numai prin procedeul calculului simplificat, atunci pentru verificarea exactităţii calculului este indicat să se schimbe originea variaţiei arbitrare, luînd o altă valoare a lui a şi calculînd încă odată media seriei.

În cazul cînd intervalele sînt neegale, cele două constante se aleg astfel încît să simplifice calculul mediei, respectiv reducerea termenilor ca şi simplificarea lor prin K trebuie să se facă separat.

6.2.1.4. MEDIA CARACTERISTICII ALTERNATIVE

Despre sistemul de caracteristici statistice s-a tratat pe larg în capitolele precedente.

Caracteristicile alternative sînt acelea care, pentru fiecare unitate, admit fie forma sa directă de manifestare, fie opusul ei. Aceasta înseamnă că unităţile colectivităţii se pot împărţi în două: unele la care se înregistrează forma directă de manifestare a caracteristicii şi o altă parte a colectivităţii la care se înregistrează opusul ei. De exemplu, un produs poate fi acceptat ca bun sau el este declarat rebut. Mai mult, orice variabilă statistică poate fi exprimată şi sub formă de caracteristică alternativă. De exemplu, fiecare muncitor se caracterizează printr-un anumit grad de îndeplinire a normelor de producţie. Deci, gradul de îndeplinire a nor-

Scanned with OKEN Scanner

melor de producţie poate fi considerat drept o variabilă nealternativă cu variaţia continuă. Dacă se împarte colectivitatea de muncitori în două subcolective : muncitori care au îndeplinit norma şi muncitori care n-au îndeplinit norma, atunci această variabilă nealternativă se transformă într-o variabilă alternativă. Sau, un alt exemplu : în cazul populaţiei poate să intereseze nu numai distribuţia populaţiei după vîrstă (care este o variabilă nealternativă) ci şi proporţia celor care se găsesc pînă la o anumită limită de vîrstă, de pildă proporţia numărului de locuitori ai unei ţări sub 25 ani şi proporţia numărului de locuitori peste 25 ani.

Rezultă, deci, că în cazul caracteristicilor alternative se întîlnesc doar două variante : unele la care se înregistrează forma directă de manifestare, deci, un răspuns afirmativ, şi altele la care se întîlneşte opusul ei şi se înregistrează, deci, un răspuns negativ. Caracteristicile alternative apar de tipul „da — nu" şi sînt destul de frecvent întîlnite în practica statistică. Pentru a calcula media acestor caracteristici trebuie să se cuantifice, adică să se exprime numeric cele două variante pe care le pot lua aceste caracteristici.

Pentru a le exprima numeric, se vor considera, în mod convenţioal, variantele cu răspuns afirmativ ca avînd valoarea 1, iar variantele cu răspuns negativ ca avînd valoarea 0. Deci, în acest caz nu sînt decît două valori ale caracteristicii : $x_1 = 1$; $x_2 = 0$, care vor diferi de la o colectivitate la alta sau de la o grupă la alta, numai prin frecvenţa lor.

Dacă se notează numărul total al unităţilor observate cu N, frecvenţa unităţilor care posedă caracteristica cu M, frecvenţa unităţilor care nu posedă caracteristica va fi $N - M$.

Pentru caracteristicile alternative, distribuţia de frecvenţă va fi cea indicată în tabelul 6.4.

Tabelul 6.4

Distribuţia de frecvenţe a caracteristicii alternative

Variantele caracteristicii alternative	Răspunsul înregistrat (variantele înregistrate)	Valoarea caracteristicii (x_i)	Frecvenţa (f_i)
— Unităţile care posedă caracteristica	da	$x_1 = 1$	$f_1 = M$
— Unităţile care nu posedă caracteristica	nu	$x_2 = 0$	$f_2 = N - M$
	—	—	$\sum_{i=1}^{2} f_i = N$

Fiind o distribuţie de frecvenţă, media se va calcula aplicînd formula mediei aritmetice ponderate :

$$\bar{x} = \frac{\sum_{i=1}^{2} x_i f_i}{\sum_{i=1}^{2} f_i} = \frac{(1 \cdot M) + O(N - M)}{M + (N - M)} = \frac{M}{N}.$$

Scanned with OKEN Scanner

Reiese că media caracteristicii alternative se obţine raportînd numărul unităţilor la care s-a înregistrat caracteristica cu răspuns afirmativ la numărul total al unităţilor. Cum unităţile colectivităţii se divid în două părţi : unele care posedă caracteristică şi altele care nu posedă caracteristică, înseamnă că, în acest caz, media caracteristicii alternative poate fi considerată şi ca mărime relativă de structură, care după cum se ştie se mai numeşte şi greutate specifică sau frecvenţă relativă.

Pentru a deosebi media caracteristicii alternative de media caracteristicii nealternative, prima se mai notează cu p.

Deci, media caracteristicii alternative este dată de formula :

$$p = \frac{M}{N} \tag{6.26}$$

Fiind o mărime medie cu caracter de mărime relativă de structură, ea se exprimă, de regulă sub formă de procente.

Se presupune, de exemplu, că într-un lot de 10 000 produse s-au înregistrat 200 rebuturi şi se cere să se afle care este media rebuturilor pe întregul lot.

Deci $N = 10\,000$; $M = 200$; media va fi :

$$p = \frac{M}{N} = \frac{200}{10\,000} = 0{,}02 \text{ sau } 2\%.$$

Aceasta înseamnă că, în medie, la fiecare 100 de produse recepţionate pot să apară cîte 2 rebuturi.

6.2.2. Media armonică simplă şi ponderată

Media armonică a termenilor unei serii se defineşte ca acea valoare a cărei mărime inversă este media aritmetică calculată din valorile inverse ale termenilor aceleiaşi serii.

Utilizînd notaţia cunoscută, $x_1, x_2, \ldots, x_n$, pentru termenii seriei, valorile lor inverse vor fi $\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_n}$. Media armonică se notează cu $\bar{x}_h$, deci, potrivit formulei :

$$\frac{1}{\bar{x}_h} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i}. \tag{6.27}$$

De aici se poate deduce formula mediei armonice simple :

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}, \tag{6.28}$$

În mod analog, media armonică ponderată se va deduce efectuîndu-se produsele de frecvenţă dintre valorile inverse ale caracteristicii şi ponde-

Scanned with OKEN Scanner

rea lor, adică va fi :

$$\bar{x}_h = \frac{\sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} \frac{1}{x_i} f_i}. \qquad (6.29)$$

Comparînd formula mediei armonice cu cea a mediei aritmetice se pot stabili anumite relaţii utile pentru folosirea în practică a celor două valori medii.

1° Dacă toate valorile caracteristicii sînt pozitive, atunci media armonică este mai mică decît media aritmetică a aceloraşi valori. Deci : $\bar{x}_h < \bar{x}_a$.

2° Dacă se consideră două variabile dependente funcţional $\left(y = \frac{1}{x}\right)$ şi pentru o variabilă se foloseşte media aritmetică, atunci pentru cealaltă variabilă este obligatoriu să se utilizeze media armonică, deoarece raportul de invers proporţionalitate se realizează şi între cele două valori medii. De exemplu, nivelul productivităţii muncii se poate exprima, în mod direct, prin cantitatea de produse ce revine pe o unitate de timp, şi indirect, prin cantitatea de timp ce revine pe unitatea de produs, deci cele două valori individuale sînt într-un raport de invers proporţionalitate.

Legătura dintre cele două căi de măsurare a productivităţii muncii se realizează prin faptul că volumul total al cheltuielilor de timp de muncă este egal cu produsul dintre volumul producţiei (q) şi consumul specific de timp de muncă pentru un produs (t). Aceasta înseamnă, în continuare, că productivitatea muncii (w) are ca ponderi cheltuielile totale de timp de muncă (T), iar consumul specific de timp de muncă (t) are ca ponderi cantităţile de produse.

În consecinţă, dacă cele două caracteristici sînt considerate dependente între ele, se pune problema să se aleagă ce fel de medie se va folosi pentru a păstra aceiaşi relaţie de invers proporţionalitate şi între mărimile medii calculate. Notînd valorile individuale ale productivităţii muncii cu (x), iar cele ale consumului specific de timp de muncă cu (y), vor apare relaţiile : $x_i = \frac{1}{y_i}$, respectiv $y_i = \frac{1}{x_i}$, pentru care se va obţine o repartiţie bidimensională cu frecvenţe comune (f_{xy}) :

Tabelul 6.5

Repartiţia productivităţii muncii şi a consumului de timp de muncă

Nivelul productivităţii muncii (x)	Consumul specific de timp de muncă (y)	Cheltuieli totale de timp de muncă f_{xy}
1	2	3
W_1	t_1	T_1
W_2	t_2	T_2
.	.	.
.	.	.
.	.	.
W_m	t_m	T_m
$\bar{W}$	$\bar{t}$	$\sum_{i=1}^{m} T_i$

Prima serie de distribuţie — a nivelurilor productivităţii muncii — este formată din datele coloanelor 1 şi 3, iar cea de-a doua serie de distribuţie — a consumurilor specifice de timp de muncă — este formată din datele coloanelor 2 şi 3. Astfel prezentate, cele două serii au aceleaşi ponderi pentru ambele caracteristici, adică numărul total de om-ore lucrate (col. 3).

Valorile individuale ale consumului specific de timp de muncă sînt invers proporţionale cu cele ale productivităţii muncii, ceea ce determină şi raportul lor faţă de ponderea aleasă. Dacă pentru calculul nivelului productivităţii muncii s-a folosit media aritmetică ponderată înseamnă că pe baza aceloraşi ponderi pentru consumul mediu de timp de muncă se va aplica media armonică ponderată:

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i f_{x_i y_i}}{\sum_{i=1}^{m} f_{x_i y_i}}; \quad \bar{y} = \frac{\sum_{i=1}^{m} f_{x_i y_i}}{\sum_{i=1}^{m} \frac{1}{y_i} f_{x_i y_i}},$$

3°. Dacă cele două variabile sînt independente între ele, atunci calculul mediei trebuie să se facă în raport cu conţinutul caracteristicii şi a distribuţiei sale de frecvenţă.

În continuare se consideră aceleaşi variabile (productivitatea muncii şi consumul de timp de muncă) ca fiind independente între ele.

Pentru a putea determina nivelul mediu al productivităţii muncii şi al consumului specific de timp de muncă pe unitatea de produs, se procedează ca în tabelele următoare.

Tabelul 6.6

Distribuţia productivităţii muncii

Nivelul productivităţii muncii (x)	Consumul total de timp de muncă f_x
W_1	T_1
W_2	T_2
⋮	⋮
W_m	T_m
$\bar{W}$	$\sum_{i=1}^{m} T_i$

Tabelul 6.7

Distribuţia consumului specific de timp de muncă

Consum specific de timp de muncă (y)	Producţia obţinută f_y
t_1	q_1
t_2	q_2
⋮	⋮
t_m	q_m
$\bar{t}$	$\sum_{i=1}^{m} q_i$

În ambele cazuri se vor folosi medii aritmetice cu ponderi diferite:

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_{x_i}}; \quad \bar{y} = \frac{\sum_{i=1}^{m} y_i f_{y_i}}{\sum_{i=1}^{m} f_{y_i}},$$

Scanned with OKEN Scanner

4° Media armonică poate să fie egală cu media aritmetică calculată din aceleaşi valori ale caracteristicii dar folosindu-se sisteme de ponderare diferite. Astfel, pentru media aritmetică se pot folosi ponderile reale ale valorilor individuale, iar pentru media armonică ponderile compuse luate sub forma produselor, adică :

$$\bar{x}_a = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i}; \quad \bar{x}_h = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} \frac{1}{x_i}(x_i f_i)}$$

Este uşor de urmărit, în formula mediei armonice, că la numitor se poate simplifica cu x şi media armonică se transformă în medie aritmetică ponderată :

$$\bar{x}_h = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} \frac{1}{x_i}(x_i f_i)} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i} = \bar{x}_a. \tag{6.30}$$

Rezultă, deci, că *există un caz particular în care media armonică apare ca formă transformată a mediei aritmetice. Media armonică este egală cu media aritmetică numai în cazul în care media armonică are ca ponderi produsele de frecvenţă xf (dacă variabila s-a notat cu x)*. Această formulă are o mare importanţă practică[1], deoarece există frecvente cazuri cînd se dispune de aceste produse de frecvenţă şi nu de ponderile reale (f).

6.2.3. Media pătratică simplă şi ponderată

Media pătratică se calculează pe baza ridicării termenilor la pătrat. Media pătratică corespunde relaţiei :

$$x_1^2 + x_2^2 + \ldots + x_n^2 = \sum_{i=1}^{n} x_i^2. \tag{6.31}$$

Aceata înseamnă că media pătratică este valoarea cu care, dacă se înlocuiesc toţi termenii luaţi la pătrat şi însumînd, se obţine aceeaşi sumă ca în cazul însumării directe a pătratelor termenilor individuali, adică :

$$\bar{x}_p^2 + \bar{x}_p^2 + \ldots + \bar{x}_p^2 = \sum_{i=1}^{n} x_i^2 \tag{6.32}$$

în care :

$\bar{x}_p$ reprezintă media pătratică,

[1] Vezi paragraful şi calculul indicelui armonic al preţurilor la capitolul „Indici".

Scanned with OKEN Scanner

$\bar{x}_p$ fiind o valoare constantă și pătratul ei este tot o mărime constantă, deci :

$$n\bar{x}_p^2 = \sum_{i=1}^{n} x_i^2 \tag{6.33}$$

de unde se poate deduce formula mediei pătratice prin extragerea rădăcinii pătrate :

$$\bar{x}_p = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} \tag{6.34}$$

În cazul unei distribuții de frecvență se vor efectua mai întîi produsul dintre fiecare termen ridicat la pătrat și ponderea sa și pe baza acestora se va calcula media pătratică ponderată :

$$\bar{x}_p = \sqrt{\frac{\sum_{i=1}^{m} x_i^2 f_i}{\sum_{i=1}^{m} f_i}} \tag{6.35}$$

Media pătratică se folosește, în special, atunci cînd se dă o mai mare importanță nivelurilor mai mari din cadrul seriei respective. Ea se folosește, în special, în demografie, în statistica industrială la calculul unor indicatori medii privind diferite mijloace de producție și, cel mai frecvent, la calculul unor indicatori de variație (abaterea medie pătratică).

Media pătratică are o serie de proprietăți, dintre care unele au importanță deosebită pentru aplicarea ei în practica statistică.

Media pătratică este întotdeauna mai mare decît media aritmetică a aceloraşi termeni, indiferent de semnul pe care îl au, deoarece prin ridicarea la pătrat toți termenii devin pozitivi.

Cu cît termenii au o valoare individuală mai ridicată, cu atît termenii respectivi vor influența în mai mare măsură asupra formării nivelului mediu. Din această cauză media pătratică se va folosi atunci cînd trebuie să se acorde o importanță mai mare nivelurilor mai ridicate din cadrul seriei pentru care se calculează media.

Aceste proprietăți, considerate ca două avantaje ale mediei pătratice, servesc cel mai bine în cazul calculării abaterii mediei, unde apar și valori negative, iar mărimea absolută a acestora influențează diferit asupra omogenității seriei.

6.2.4. Media geometrică simplă și ponderată

Spre deosebire de celelalte medii prezentate pînă aici, care se bazează pe relații de însumare între termenii seriei, media geometrică se bazează pe relația de produs dintre ei.

Scanned with OKEN Scanner

Media geometrică reprezintă acea valoare cu care, dacă se înlocuiesc toţi termenii seriei şi se face produsul lor, valoarea la care se ajunge este egală cu produsul termenilor reali, adică :

$$x_1 . x_2 . x_3 \ldots x_n = \prod_{i=1}^{n} x_i \tag{6.36}$$

$$\bar{x}_g . \bar{x}_g . \bar{x}_g \ldots \bar{x}_g = \prod_{i=1}^{n} x_i \tag{6.37}$$

în care :

Π este semnul produsului.

Şi aici $\bar{x}_g$ este tot o valoare constantă, deci produsul ei va fi egal cu valoarea sa ridicată la o putere egală cu numărul termenilor pe care îi substituie, deci :

$$\bar{x}_g^n = \prod_{i=1}^{n} x_i \tag{6.38}$$

De aici se deduce formula mediei geometrice :

$$\bar{x}_g = \sqrt{\prod_{i=1}^{n} x_i} \tag{6.39}$$

În cazul unei serii de distribuţii de frecvenţe, fiecare termen trebuie să fie luat în funcţie de frecvenţa sa. Aceasta înseamnă că, în cazul mediei geometrice, fiecare termen se înmulţeşte de un număr egal cu frecvenţa lui în cadrul seriei, deci frecvenţele devin puterile la care se ridică fiecare termen :

$$x_1^{f_1} . x_2^{f_2} \ldots x_m^{f_m} = \prod_{i=1}^{m} x_i^{f_i} \tag{6.40}$$

$$\bar{x}_g^{f_1} . \bar{x}_g^{f_2} \ldots \bar{x}_g^{f_m} = \prod_{i=1}^{m} x_i^{f_i} \tag{6.41}$$

Relaţia a doua se poate restrînge :

$$(\bar{x}_g)^{\sum_{i=1}^{m} f_i} = \prod_{i=1}^{m} x_i^{f_i} \tag{6.42}$$

de unde formula mediei geometrice se poate scrie :

$$\bar{x}_g = \sqrt[\sum_{1}^{m} f_i]{\prod_{i=1}^{m} x_i^{f_i}} \tag{6.43}$$

Scanned with OKEN Scanner

1° Se poate observa că, în cazul mediei geometrice, este suficient ca un singur termen al seriei să fie egal cu zero, pentru ca media geometrică să nu se poată calcula, iar în cazul cînd unul din termeni este negativ media să devină imaginară (cazuri fără sens).

2° În cazul mediei geometrice, abaterile termenilor seriei faţă de medie nu se mai calculează sub formă de diferenţe, ci sub formă de rapoarte $\left(\frac{x_i}{\bar{x}_g}\right)$, iar produsul acestora este egal cu 1. Această proprietate se poate scrie astfel:

$$\prod_{i=1}^{n}\left(\frac{x_i}{\bar{x}_g}\right) = 1. \tag{6.44}$$

Într-adevăr, scriind membrul I dezvoltat, se obţine:

$$\frac{x_1}{\bar{x}_g}\cdot\frac{x_2}{\bar{x}_g}\cdots\frac{x_n}{\bar{x}_g} = \frac{\prod_{i=1}^{n} x_i}{(\bar{x}_g)^n} = \frac{\prod_{i=1}^{n} x_i}{\prod_{i=1}^{n} x_i} = 1. \tag{6.45}$$

Pentru a putea fi calculată atît media geometrică simplă, cît şi cea ponderată trebuie să se logaritmeze. Prin logaritmare se va obţine:

— pentru media geometrică simplă:

$$\log \bar{x}_g = \frac{\sum_{i=1}^{n} \log. x_i}{n}. \tag{6.46}$$

— pentru media geometrică ponderată:

$$\log. \bar{x}_g = \frac{\sum_{i=1}^{m} (f_i . \log. x_i)}{\sum_{i=1}^{n} f_i} \tag{6.47}$$

3° Deci, prin aplicarea logaritmilor media geometrică se transformă într-o medie aritmetică a logaritmilor factorilor iar antilogaritmul ei este o valoare mai mică decît media aritmetică calculată din valorile reale ale termenilor seriei.

Media geometrică se foloseşte cel mai frecvent în cazul seriilor dinamice, la calculul mediilor din mărimile relative ale dinamicii, între care există o relaţie de produs.

În cazul seriilor de distribuţie de frecvenţe, media geometrică se foloseşte mai rar. Ea este indicată să se folosească atunci cînd seria prezintă variaţii foarte mari între termeni sau un pronunţat caracter de asime-

Scanned with OKEN Scanner

trie. Prin logaritmare abaterile dintre termenii seriei se micşorează şi se obţine un grad mai mare de concentrare a frecvenţelor.

Date fiind însă dificultăţile de calcul al mediei geometrice, ea se foloseşte foarte rar în analiza seriilor de distribuţie, unde, de regulă, se foloseşte media aritmetică.

Între mediile prezentate pînă aici se poate stabili următoarea relaţie de ordine:

$$\bar{x}_h < \bar{x}_g < \bar{x}_a < \bar{x}_p \qquad (6.48)$$

6.2.5. Calculul mediei unor indicatori derivaţi, exprimaţi prin mărimi relative parţiale sau mărimi medii parţiale

În statistica social-economică se întîlnesc frecvent caracteristici secundare (în sensul că se definesc în procesul de prelucrare) determinate prin indicatori derivaţi sub forma de mărimi relative parţiale sau mărimi medii parţiale şi care sînt interpretate ca variabile aleatoare.

Mărimile relative parţiale, fiind variante ale unor variabile aleatoare, şi pentru ele se stabilesc serii de repartiţii de frecvenţe, deoarece şi ele pot lua valori de la x_1 la x_m, astfel:

$$x_1 = \frac{y_1}{z_1}; \quad x_2 = \frac{y_2}{z_2}; \ldots; \quad x_m = \frac{y_m}{z_m}. \qquad (6.49)$$

Cei doi indicatori (y, z) sînt variabilele unor caracteristici primare, exprimate, de regulă, sub formă de mărimi absolute.

Pentru a calcula media acestor mărimi relative parţiale este necesar să se identifice care este ponderea cu care ele intervin în calcul, atunci cînd se referă la întreaga colectivitate.

După cum se ştie, mărimea unui raport depinde, în primul rînd, de baza de raportare, deci se poate considera că numitorii parţiali $(z_{1,2} \ldots, m)$ sînt ponderile caracteristicii derivate.

Dacă (x_i) este valoarea fiecărui raport, iar numitorii parţiali (z_i) sînt ponderile (f_i) acestei variabile rezultă că numărătorii parţiali (y_i) sînt egali cu produsele de frecvenţă $(x_i f_i)$, adică:

$$y_i = x_i f_i, \quad \text{de unde:} \quad x_i = \frac{y_i}{z_i} = \frac{x_i f_i}{f_i}.$$

În practica statistică se pot întîlni două situaţii diferite, corespunzînd la două variante de calcul al mediei.

1° Se cunosc valorile parţiale ale mărimilor relative (x_i) şi numitorii parţiali din care s-au calculat (z_i). Dacă $z_i = f_i$, înseamnă că se va putea calcula o medie aritmetică ponderată deoarece se cunosc atît valorile individuale ale variabilei derivate (x_i), cît şi frecvenţele lor de apariţie (f_i),

Scanned with OKEN Scanner

adică :

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i}. \qquad (6.50)$$

2° Se cunosc valorile parțiale ale mărimilor relative (x_i) și numărătorii parțiali din care s-au calculat (y_i). Dacă $y_i = x_i f_i$, înseamnă că se poate aplica o medie armonică cu frecvență compusă ($x_i f_i$) aplicată valorilor individuale ale variabilei secundare (x_i), adică :

$$\bar{x}_h = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} \frac{1}{x_i} x_i f_i}. \qquad (6.51)$$

Se poate demonstra cu ușurință că cele două variante de calcul conduc la aceeași valoare. Astfel, dacă se fac simplificările cu (x) la numitor, se obține :

$$\bar{x}_h = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} \frac{1}{x_i} x_i f_i} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i} = \bar{x}_a. \qquad (6.52)$$

Rezultă că, pentru a calcula media unor mărimi relative parțiale, nu este suficient să se cunoască numai acestea. Trebuie, în plus, să se cunoască fie numitorii, fie numărătorii parțiali din care au fost calculate, pentru a putea ține seama de frecvența cu care ele apar în distribuția de valori din care fac parte.

Se poate calcula o medie simplă a lor numai dacă se știe că toate mărimile relative parțiale au fie numărătorii, fie numitorii egali — cazuri foarte rare în cadrul fenomenelor sociale și economice.

În analiza fenomenelor social-economice, colectivitățile studiate sînt împărțite în grupe, caracterizate prin medii parțiale (pe fiecare grupă) și dacă colectivitatea este omogenă, mediile parțiale se sintetizează într-o medie generală.

Mai mult, analiza statistico-economică trebuie să țină seama de faptul că fenomenele pe care le studiază statistica sînt constituite, în mod permanent, în unități complexe care sînt formate din mai multe unități simple și fiecare unitate complexă trebuie să fie caracterizată prin anumite valori medii, care, în raport cu media pe întreaga colectivitate, pot fi considerate ca medii parțiale.

La calculul mediei generale din medii parțiale se pot întîlni două situații :

1°) grupele pentru care s-au calculat mediile parțiale sînt egale între ele din punctul de vedere al numărului de unități ;

Scanned with OKEN Scanner

2°) numărul de unităţi diferă de la o grupă la alta.

În primul caz, frecvenţele grupelor fiind egale înseamnă că media generală se va calcula ca medie aritmetică simplă a mediilor parţiale, adică :

$$\bar{x}_0 = \frac{\sum_{i=1}^{r} \bar{x}_i}{r}, \tag{6.53}$$

în care :

$\bar{x}_0$ reprezintă media colectivităţii generale ;
$\bar{x}_i$ — media de grupă ;
r — numărul de grupe în care a fost împărţită colectivitatea.

În cea de-a doua situaţie, cînd numărul de unităţi diferă de la o grupă la alta, media se va calcula ca o medie aritmetică ponderată de forma :

$$\bar{x}_0 = \frac{\sum_{i=1}^{r} \bar{x}_i n_i}{\sum_{i=1}^{r} n_i}, \tag{6.54}$$

în care :

n_i reprezintă frecvenţele cumulate (f) pentru fiecare grupă.

În acest caz mediile de grupă ($\bar{x}_{1,2} \ldots r$) se vor folosi ca şi cum ar fi valorile individuale ale caracteristicii, iar ponderea fiecărei medii va fi suma frecvenţelor din grupa respectivă.

Şi aici poate apare o a doua variantă şi anume, cînd nu se cunosc frecvenţele de la fiecare grupă, ci numai produsele de frecvenţă ($\bar{x}_i n_i$). Deci, şi aici se poate folosi media armonică în cazul ponderilor compuse sub forma produselor de frecvenţă :

$$\bar{x}_h = \frac{\sum_{i=1}^{r} \bar{x}_i n_i}{\sum_{i=1}^{r} \frac{1}{\bar{x}_i} \bar{x}_i n_i}. \tag{6.55}$$

La calculul *mediei generale din medii parţiale a unei caracteristici alternative* trebuie să se ţină seama că această variabilă este de tipul unei mărimi relative şi că, pentru calculul mediei generale, se pot folosi următoarele două variante :

Media generală a unei caracteristici alternative calculată ca medie aritmetică ponderată :

$$\bar{p} = \frac{\sum_{i=1}^{r} p_i N_i}{\sum_{i=1}^{r} N_i}, \tag{6.56}$$

în care:

$\bar{p}$ reprezintă media generală a unei caracteristici alternative;

p_i — media de grupă a caracteristicii alternative;

N_i — frecvența unităților din fiecare grupă.

6° Media generală a unei caracteristici alternative calculată ca medie armonică ponderată:

$$\bar{p} = \frac{\sum_{i=1}^{r} p_i N_i}{\sum_{i=1}^{r} \frac{1}{p_i} p_i N_i} = \frac{\sum_{i=1}^{r} M_i}{\sum_{i=1}^{r} \frac{1}{p_i} M_i}, \tag{6.57}$$

în care:

$p_i N_i = M_i$, deoarece $p_i = \frac{M_i}{N_i}$, deci se înlocuiește în formulă p_i cu echivalentul său și se va obține:

$$\bar{p} = \frac{\sum_{i=1}^{r} M_i}{\sum_{i=1}^{r} \frac{1}{\frac{M_i}{N_i}} p_i N_i} = \frac{\sum_{i=1}^{r} M_i}{\sum_{i=1}^{r} \frac{M_i}{N_i} N_i} = \frac{\sum_{i=1}^{r} M_i}{\sum_{i=1}^{r} N_i}. \tag{6.58}$$

Rezultă că, în cazul caracteristicii alternative, media generală se poate calcula după una din cele două formule prezentate în funcție de datele care se cunosc. Exemple de calcul al mediei generale pe baza mediilor parțiale vor fi prezentate la calculul indicatorilor de variație.

6.2.6. Valori medii de poziție sau de structură

Pentru completarea analizei seriilor de distribuție este necesar să se calculeze și anumite valori medii de poziție sau medii de structură, printre care mediana și modulul sînt cele mai frecvent utilizate.

6.2.6.1. MEDIANA

Mediana (*Me*) reprezintă valoarea centrală a unei serii statistice, ordonate în mod crescător sau descrescător și care împarte termenii seriei în două părți egale: 50% din numărul termenilor vor fi mai mici decît valoarea medianei și 50% se vor plasa peste valoarea ei.

Locul medianei într-o serie de n termeni este dat de formula:

$$\text{locul } Me = \frac{n+1}{2}. \tag{6.59}$$

Scanned with OKEN Scanner

În seriile cu termeni impari, mediana este determinată de termenul empiric, care împarte seria în două. În seriile cu număr par de termeni, mediana se determină ca medie aritmetică simplă din valoarea celor doi termeni centrali.

Mai dificil de calculat este valoarea medianei pentru o serie de distribuţie de frecvenţe de intervale, unde se pune problema să se găsească atît intervalul median, cît şi valoarea medianei.

Locul medianei într-o serie de distribuţie de frecvenţă, în care $n = \sum_{i=1}^{m} f_i$, se determină după formula:

$$\text{locul } Me = \frac{\left(\sum_{i=1}^{m} f_i\right) + 1}{2}. \qquad (6.60)$$

În cazul în care suma frecvenţelor este foarte mare (peste 500 unităţi) se poate renunţa la +1, deci locul medianei este:

$$\frac{\sum_{i=1}^{m} f_i}{2} = \frac{1}{2} \sum_{i=1}^{m} f_i. \qquad (6.61)$$

În continuare se va calcula valoarea medianei pentru seria de distribuţie folosită la determinarea mediei aritmetice ponderate. Pentru a putea trece la calculul medianei este necesar să se cumuleze frecvenţele, care se vor înscrie în coloanele tabelului de calcul (vezi tabelul 6.8).

Seria avînd frecvenţe peste 500, locul medianei va fi:

$$\text{locul } Me = \frac{\sum_{i=1}^{m} f_i}{2} = \frac{600}{2} = 300{,}0.$$

Tabelul 6.8

Repartiţia muncitorilor după mărimea retribuţiei lunare

Grupe de muncitori după retribuţia obţinută	Număr de muncitori (f_i)	Frecvenţa cumulată (f_c)
A	1	2
sub 1 500	48	48
1 500—1 800	72	120
1 800—2 100	108	228
2 100—2 400	123	351
2 400—2 700	105	456
2 700—3 000	84	540
peste 3 000	60	600
Total	600	

Scanned with OKEN Scanner

Intervalul median va fi considerat acela în care frecvenţele cumulate depăşesc locul medianei în serie. În exemplul luat, mediana se găseşte între unitatea cu numărul de ordine 300 şi 301. Urmărind în tabel aceste două unităţi se observă că ele nu se găsesc nici în primul interval care conţine numai 48 de muncitori, nici în următoarele două care conţin împreună 120 şi, respectiv 228, ci după al patrulea interval în care s-au cuprins în total 351 de unităţi. Intervalul median este grupa în care muncitorii au o retribuţie între 2 100 şi 2 400 lei. Deci :

$$2\,100 < Me < 2\,400.$$

Pentru a trece la calculul valorii medianei trebuie precizat locul medianei în intervalul median, ceea ce înseamnă că din numărul total al frecvenţelor celor patru intervale să se elimine cele care nu aparţin intervalului median.

Astfel, locul medianei în intervalul median va fi :

$$\frac{1}{2}\sum_{i=1}^{m} f_i - \Sigma f_p, \tag{6.62}$$

în care :

Σf_p reprezintă frecvenţele intervalelor precedente intervalului median.

Pentru a afla valoarea medianei, se ia în considerare limita inferioară a intervalului median la care se adaugă o parte din mărimea intervalului proporţională cu raportul dintre frecvenţe. Intervalul median corespunde pentru toţi muncitorii, adică în intervalul de variaţie egal cu 300 lei sînt 123 de muncitori, dintre care primii 72 de muncitori au retribuţii sub valoarea retribuţiei mediane. Presupunînd că retribuţiile individuale s-ar distribui uniform în interval, formula de calcul al medianei va fi :

$$Me = x_0 + d\frac{\frac{1}{2}(\Sigma f + 1) - \Sigma f_p}{f_m}, \tag{6.63}$$

în care :

x_0 reprezintă limita inferioară a intervalului median ;
d — mărimea intervalului median ;
f_m — frecvenţa intervalului median.

Aplicînd formula în exemplul luat se obţine :

$$Me = 2\,100 + 300\,\frac{\frac{600}{2} - 228}{123} = 2\,272{,}80 \text{ lei.}$$

Valoarea mediană a acestei serii este de 2 272,80 lei, rezultînd astfel că 300 de muncitori au retribuţii ce se găsesc sub această valoare şi 300 muncitori au o retribuţie mai mare decît aceasta. În raport cu valoarea medie, mediana este foarte apropiată, reprezentînd o abatere de numai 5,70 lei.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

6.2.6.2. MODULUL

Modulul (Mo) reprezintă valoarea caracteristicii care are frecvența maximă. Dacă se reprezintă grafic seria, atunci modulul corespunde termenului cu ordonata maximă. Rezultă, deci, că valoarea modulului nu se întîlnește decît într-o serie de distribuție de frecvențe. Cînd într-un poligon de frecvențe se întîlnesc mai multe ordonate maxime se spune că este o serie plurimodală.

Într-o serie de distribuție pe variante valoarea modulului se determină direct.

Pentru o serie de distribuție pe intervale valoarea modulului trebuie calculată. Intervalul modal se consideră intervalul care are frecvența cea mai mare.

În seria prezentată pentru calculul medianei, intervalul modal este cel cuprins între 2 100—2 400 lei, deoarece aici frecvența este maximă (123). Rezultă că pentru distribuția muncitorilor din întreprinderea respectivă : $2\,100 < Mo < 2\,400$. Dar 2 100 este limita superioară pentru intervalul precedent, iar 2 400 este limita inferioară pentru cel următor. Deci, la calculul modulului trebuie să se țină seamă și de frecvențele intervalelor alăturate.

Formula de calcul al modulului, într-o serie de intervale este :

$$Mo = x_0 + d\frac{\Delta_1}{\Delta_1 + \Delta_2} \qquad (6.64)$$

în care :

x_0 reprezintă limita inferioară a intervalului modal ;
d — mărimea intervalului modal ;
Δ_1 — $f_m - f_{m-1}$;
Δ_2 — $f_m - f_{m+1}$;
f_m — frecvența intervalului modal ;
f_{m-1} — frecvența intervalului precedent celui modal ;
f_{m+1} — frecvența intervalului următor celui modal.

În seria prezentată :

$$Mo = 2\,100 + 300\frac{123-108}{(123-108)+(123-105)} = 2\,236{,}36 \text{ lei.}$$

Modulul reprezintă, deci, o abatere mai mare față de medie (2 278,50 — —2 236,35 = 42,14 lei). Fiind pozitivă, această abatere arată că frecvențele termenilor mai mici sînt mai numeroase, în comparație cu frecvențele termenilor mai mari.

Pe grafic, valoarea modală se determină folosind histograma sau poligonul frecvențelor. Modulul este valoarea de pe abscisă căreia îi corespunde ordonata maximă (fig. 6.1).

Deci, modulul este una din valorile care definește forma de distribuție a unei caracteristici, exprimînd prin mărimea sa nivelul cel mai frecvent întîlnit într-o perioadă de timp, de exemplu, retribuția, productivitatea muncii și categoria de calificare cel mai frecvent întîlnită într-o ramură etc.

Uneori modulul poate înlocui mărimea medie. De exemplu, în producţia şi circulaţia de bunuri de consum, prin anumite sondaje se determină care este talia cea mai frecventă la confecţii, numărul cel mai solicitat la pantofi etc. Pentru astfel de produse media nu are sens să fie calculată.

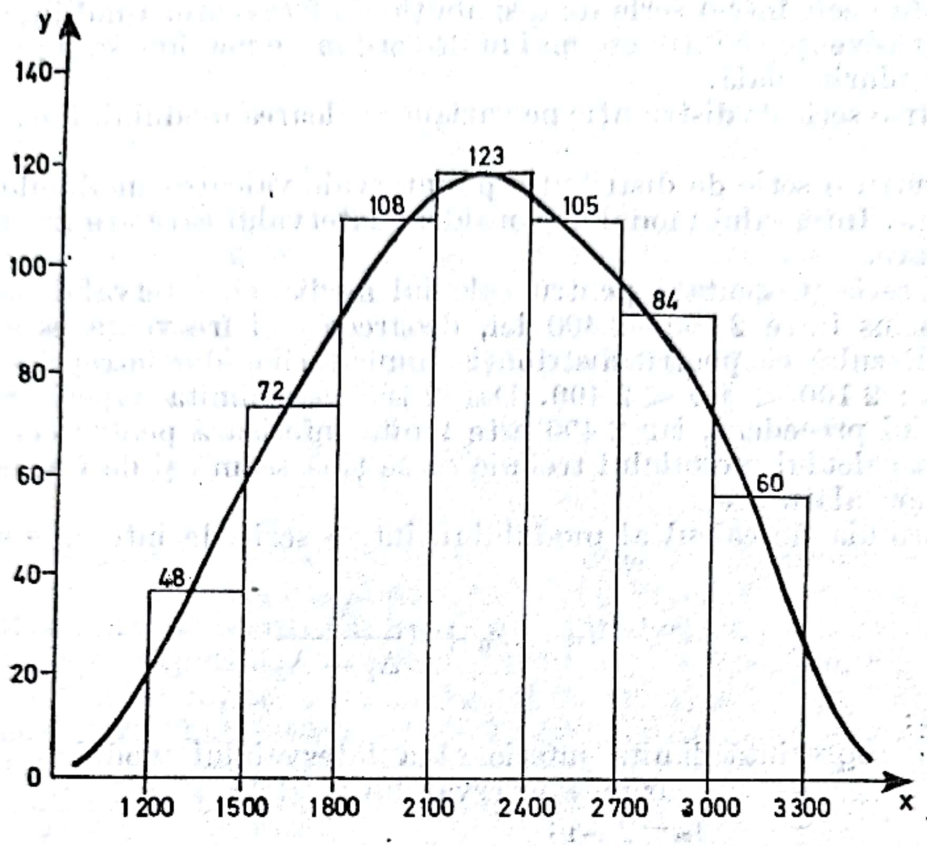


Fig. 6.1. Repartiţia muncitorilor după mărimea retribuţiilor lunare.

Mediana şi modulul se exprimă în aceleaşi unităţi de măsură ca şi variabila pentru care se determină şi prezintă avantajul că valoarea lor nu se modifică dacă se închid în mod convenţional intervalele deschise. Bazîndu-se numai pe relaţiile între frecvenţe, valorile lor nu depind de valorile extreme ale caracteristicii ; în cazul mediei care sintetizează toate valorile individuale, închiderea intervalelor, în mod convenţional, în funcţie de mărimea intervalelor alăturate, influenţează asupra valorii ei.

Rezultă că media, mediana şi modulul pot caracteriza forma de variaţie a caracteristicii. În cazul cînd o distribuţie este perfect simetrică, media, mediana şi modulul coincid în valoare. Astfel de situaţii se întîlnesc foarte rar în cadrul fenomenelor social-economice, care, de regulă, prezintă o anumită tendinţă de asimetrie şi care poate fi măsurată prin coeficienţii de asimetrie.

6.2.6.3. CUARTILE, DECILE

Pentru seriile de distribuţie cu tendinţă pronunţată de asimetrie, caracterizate printr-o amplitudine mare a variaţiei, se calculează şi alţi indicatori de poziţie, cum sînt : *cuartilele*, *decilele* etc.

Scanned with OKEN Scanner

Cuartilele sînt acele valori ale caracteristicii, care separă seria în patru părţi egale :

— cuartila inferioară, notată cu Q_1 este mai mare decît 25% din termenii seriei şi mai mică decît 75% dintre ei ;

— cuartila a doua (Q_2), egală cu mediana, separă seria în două părţi : 50% din numărul unităţilor sînt mai mici decît valoarea sa şi 50% peste valoarea sa ;

— cuartila superioară (Q_3) este mai mare decît 75% din numărul termenilor şi mai mică decît 25% din numărul lor.

În mod analog, formulele de calcul ale cuartilelor vor fi :

$$Q_1 = x_0 + d\frac{\frac{1}{4}(\Sigma f + 1) - \Sigma f_p}{f_{Q_1}} \tag{6.65}$$

$$Q_2 = x_0 + d\frac{\frac{2}{4}(\Sigma f + 1) - \Sigma f_p}{f_{Q_2}} = Me \tag{6.66}$$

$$Q_3 = x_0 + d\frac{\frac{3}{4}(\Sigma f + 1) - \Sigma f_p}{f_{Q_3}}, \tag{6.67}$$

în care :

Σf_p — suma frecvenţelor intervalelor precedente locului pe care îl ocupă cuartila respectivă în serie ;

$f_{Q_1}, f_{Q_2}, f_{Q_3}$ — frecvenţele intervalelor care conţin cuartilele respective.

În cazul în care se calculează decilele, seria se divide în zece părţi egale, prima fiind decila :

$$D_1 = x_0 + d\frac{\frac{1}{10}(\Sigma f + 1) - \Sigma f_p}{f_{D_1}} \tag{6.68}$$

$$\vdots$$

$$D_9 = x_0 + d\frac{\frac{9}{10}(\Sigma f + 1) - \Sigma f_p}{f_{D_9}}, \tag{6.69}$$

în care :

Σf_p reprezintă suma frecvenţelor cumulate ale intervalelor precedente intervalului care conţine decila respectivă ;

$f_{D_1}, f_{D_2} \ldots f_{D_9}$ — frecvenţele intervalelor care cuprind decilele respective.

Scanned with OKEN Scanner

Dacă seria este cu un grad foarte mare de asimetrie, se pot calcula şi centilele. *Centilele* separă seria în 100 părţi egale, deci se pot calcula 99 de centile, adaptîndu-se în mod corespunzător formula medianei la calculul acestora.

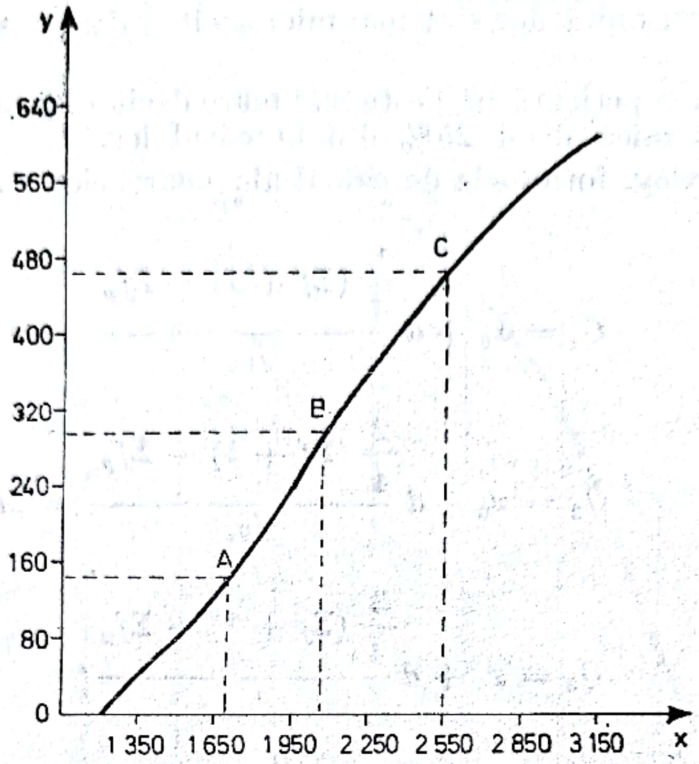


Fig. 6.2. Distribuţia personalului după retribuţia obţinută

Cuartilele, decilele, centilele se pot determina şi grafic folosind diagrama cumulativă a frecvenţelor şi reprezintă punctele corespunzătoare pe axa *Ox*.

Folosind ca elemente de bază mediile calculate din totalitatea valorilor individuale, precum şi mediile de structură, se pot calcula indicatorii variaţiei şi asimetriei, prin interpretarea cărora se va efectua analiza tipului de distribuţie către care tinde seria de date empirice.

6.3. INDICATORII VARIAŢIEI ŞI ASIMETRIEI

Practica statistică demonstrează că pentru aceeaşi colectivitate, caracteristicile care o definesc pot prezenta grade şi forme de variaţie, ce pot fi analizate la nivele diferite, după cum au fost folosite unităţi de observare simple sau unităţi de observare complexe, pentru întreaga colectivitate sau numai pentru o parte a acesteia. Apare evident faptul că, cu cît colectivitatea este mai puţin omogenă, cu atît caracterizarea ei sub raport statistic trebuie să se bazeze pe un sistem mai complex de relaţii în care sînt folosiţi indicatorii totalizatori, cît şi indicatori derivaţi. Printre

Scanned with OKEN Scanner

indicatorii derivaţi, media aritmetică calculată din totalitatea cazurilor observate reprezintă indicatorul cu care se estimează cel mai frecvent valoarea tipică ce exprimă tendinţa centrală de manifestare a unei variabile statistice, pentru care abaterile de nivel sînt considerate ca fiind rezultatul unor factori întîmplători.

Cu cît fenomenele sînt mai complexe, deci dependente de mai mulţi factori, cu atît variaţia este mai mare şi folosirea corectă a valorilor medii implică verificări riguroase cu privire la stabilitatea şi reprezentativitatea ei în funcţie de datele obţinute din observarea statistică.

După cum s-a mai arătat, media nu este o valoare reprezentativă decît pentru cazul în care ea este calculată din mărimi omogene între ele.

Pentru a verifica gradul de omogenitate al caracteristicilor pentru care se determină media este necesar să se calculeze indicatorii de variaţie de asimetrie şi exces. Aceşti indicatori trebuie să servească la:

— verificarea reprezentativităţii mediei ca valoare tipică a unei serii de date statistice;

— verificarea gradului de omogenitate a seriei;

— caracterizarea statistică a formei şi gradului de variaţie a unei caracteristici;

— compararea în timp şi spaţiu a mai multor serii statistice de distribuţie pentru aceeaşi caracteristică sau pentru caracteristici interdependente;

— cunoaşterea gradului de influenţă a factorilor după care s-a făcut gruparea unităţilor observate.

Avînd acest sistem de indicatori se pot caracteriza statistic fenomenele studiate, frecvenţa lor de apariţie, legătura dintre datele individuale şi indicatorii derivaţi calculaţi ca valori sintetice. Se poate spune, deci, că dacă prin medie se realizează o sinteză a fenomenelor variabile, prin calcularea indicatorilor de variaţie şi asimetrie se efectuează analiza acestora, prin compararea formelor concrete de manifestare cu valorile tipice, care exprimă tocmai legitatea lor de dezvoltare.

Indicatorii de variaţie permit, deci, separarea modului de acţiune a factorilor esenţiali de acţiunea factorilor întîmplători, ajutînd la identificarea felului în care acţionează factorii esenţiali de la o grupă la alta. Această funcţie de cunoaştere este folosită în special pentru acele fenomene la care, pe lîngă acţiunea separată a factorilor, se mai adaugă şi interacţiunea dintre factori.

În studiul gradului de influenţă a diferiţilor factori se folosesc indicatori simpli şi sintetici ai variaţiei şi ai analizei dispersionale în care se aplică, pe rînd, substituirea în lanţ a factorilor de influenţă. Aceasta impune în mod necesar folosirea metodei grupării unităţilor observate în funcţie de una sau mai multe caracteristici de grupare, care pot fi considerate ca factori de influenţă ai variaţiei caracteristicii ce se studiază. În acelaşi timp, folosind sistemul de indicatori ai variaţiei şi asimetriei se poate verifica însuşi modul în care s-a făcut gruparea datelor şi se poate aprecia dacă este necesar să se refacă gruparea, schimbînd numărul de grupe şi intervalul de grupare.

Se poate spune că indicatorii de variaţie pot fi folosiţi ca teste de verificare a modului de sistematizare a datelor cu ajutorul grupării.

Scanned with OKEN Scanner

n concluzie, indicatorii variației și asimetriei pot fi folosiți la caracterizarea independentă a fenomenelor, la estimarea erorilor de selecție, în analiza corelației statistice, și în general, în toate cazurile cînd se folosesc mărimile medii și trebuie să se interpreteze măsura în care ele sînt reprezentative pentru toți termenii individuali din care au fost calculate.

Indicatorii variației pot fi calculați ca indicatori simpli și ca indicatori sintetici.

6.3.1. Indicatorii simpli ai variației

În studiul variației fenomenelor interesează atît gradul de împrăștiere a valorilor înregistrate pentru o caracteristică statistică, cît și forma în care se distribuie frecvențele în funcție de aceste valori individuale.

Indicatorii simpli ai variației servesc tocmai pentru a caracteriza gradul de împrăștiere a unităților purtătoare ale caracteristicilor înregistrate. Ei se calculează pentru a măsura amplitudinea variației și abaterilor valorilor individuale de la media lor. Acești indicatori se pot exprima atît în mărimi absolute — folosind aceleași unități de măsură ca și pentru caracteristica studiată — cît și în mărimi relative, calculate în raport cu valoarea medie.

Amplitudinea absolută a variației (A) se calculează ca diferență între nivelul maxim (x_{max}) și nivelul minim (x_{min}) al caracteristicii:

$$A = x_{max} - x_{min}. \tag{6.70}$$

În cazul unei serii de distribuție de frecvențe amplitudinea variației se calculează ca diferență între limita maximă a intervalului superior și limita inferioară a intervalului inferior. Dacă intervalele sînt deschise, atunci amplitudinea variației se determină după ce s-au închis, în mod convențional, intervalele extreme.

Fiind legată de unitatea de măsură a caracteristicii, amplitudinea variației nu poate fi folosită ca măsură a variației decît pentru seriile care se referă la aceeași caracteristică înregistrată pentru aceeași unitate de timp, dar în unități de spațiu diferite, sau pentru aceeași unitate de spațiu dar pentru perioade de timp diferite.

Amplitudinea relativă a variației ($A\%$) se exprimă, de regulă, în procente și se calculează ca raport între amplitudinea absolută a variației și nivelul mediu al caracteristicii:

$$A\% = \frac{A}{\bar{x}} \times 100. \tag{6.71}$$

Amplitudinea variației nu este însă un indicator suficient de semnificativ, deoarece nu ține seama decît de valorile extreme ale caracteristicii. Or, asupra variației unui fenomen influențează toate valorile individuale și frecvențele lor de apariție. Uneori se întîlnesc cazuri cînd valorile extreme sînt foarte depărtate de celelalte valori intermediare și, în acest caz, a măsura variația numai în funcție de amplitudinea ei constituie o greșeală.

Scanned with OKEN Scanner

Amplitudinea variației se folosește în prelucrarea statistică la alegerea numărului de grupe și a mărimii intervalului de grupare. În controlul calității produselor, amplitudinea variației este un indicator care se interpretează în raport cu limitele de toleranță admise.

Abaterile individuale absolute (d) se calculează ca diferențe între fiecare variantă înregistrată și media aritmetică a acestora. Astfel:

$$d_i = x_i - \bar{x}.$$

Abaterile individuale relative ($d\%$) se calculează raportînd abaterile absolute la nivelul mediu al caracteristicii:

$$d_i = \frac{d_i}{\bar{x}} \times 100 = \frac{x_i - \bar{x}}{\bar{x}} \times 100. \qquad (6.72)$$

În analiza variației interesează, în special, abaterile maxime într-un sens sau altul:

— în mărimi absolute:

$$d_{\text{max negativă}} = x_{\min} - \bar{x}; \qquad d_{\text{max pozitivă}} = x_{\max} - \bar{x}; \qquad (6.73)$$

— în mărimi relative:

$$d_{\text{max.negativă}}\% = \frac{x_{\min} - \bar{x}}{\bar{x}} \times 100; \quad d_{\text{max.pozitivă}}\% = \frac{x_{\max} - \bar{x}}{\bar{x}} \times 100. \qquad (6.74)$$

Gradul de variație al unei caracteristici nu depinde însă numai de abaterile maxime care apar într-un sens sau altul, ci de toate abaterile variantelor înregistrate și de frecvențele lor de apariție. Numai în cazul distribuției perfect simetrice abaterea absolută maximă pozitivă coincide cu abaterea maximă negativă, iar în interiorul seriei abaterilor egale în mărimi absolute le corespund frecvențe egale de apariție, ceea ce asigură o compensare a abaterilor nu numai pe total, ci și la nivelul unităților observate.

În concluzie, indicatorii simpli ai variației, fiind calculați pe baza relațiilor dintre doi termeni ai seriei sau între fiecare termen și media lor, nu pot exprima întreaga variație a unei caracteristici înregistrate. De aceea este necesar să se calculeze și indicatorii sintetici ai variației care iau în considerație toate abaterile caracteristicii.

6.3.2. Indicatorii sintetici ai variației

Pentru a sintetiza într-o singură expresie numerică întreaga variație a unei caracteristici trebuie să se recurgă tot la o valoare medie calculată din abaterile individuale ale variantelor de la media lor.

Indicatorii sintetici ai variației sînt: abaterea medie liniară, abaterea medie pătratică, dispersia și coeficientul de variație.

Scanned with OKEN Scanner

Abaterea medie liniară ($\bar{d}$) se calculează ca o medie aritmetică simplă sau ponderată din abaterile termenilor seriei de la media lor, luate în valoare absolută.

Abaterile termenilor seriei de la media lor aritmetică se iau în valoare absolută, deoarece suma algebrică a tuturor abaterilor este egală cu zero.

Formulele de calcul ale abaterii medii liniare sînt :

— pentru o serie simplă :

$$\bar{d} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|; \tag{6.75}$$

— pentru o serie de frecvenţe absolute :

$$\bar{d} = \frac{\sum_{i=1}^{m} |x_i - \bar{x}| f_i}{\sum_{i=1}^{m} f_i} = \frac{1}{\sum_{i=1}^{m} f_i} \sum_{i=1}^{m} |x_i - \bar{x}| f_i; \tag{6.76}$$

— pentru o serie cu frecvenţe relative, exprimate în procente :

$$\bar{d} = \frac{\sum_{i=1}^{m} |x_i - \bar{x}| f_i^* \%}{100} = \frac{1}{100} \sum_{i=1}^{m} |x_i - \bar{x}| f_i^* \%. \tag{6.77}$$

Dacă frecvenţele relative sînt exprimate sub formă de coeficienţi, atunci abaterea medie liniară va fi :

$$\bar{d} = \sum_{i=1}^{m} |x_i - \bar{x}| f_i^*. \tag{6.78}$$

Abaterea medie liniară prezintă dezavantajul că nu ţine seama de faptul că abaterile mai mari în valoare absolută influenţează în mai mare măsură gradul de variaţie a unei caracteristici, în comparaţie cu abaterile mai mici. În acelaşi timp, din punct de vedere algebric, nu este indicat să se renunţe în mod arbitrar la semnul valorilor din care se calculează o valoare medie. Din aceste considerente se foloseşte ca principal indicator sintetic al variaţiei abaterea medie pătratică.

Abaterea medie pătratică sau abaterea *standard*, notată cu σ, se calculează ca o medie pătratică din abaterile tuturor variantelor seriei de la media lor aritmetică :

— pentru o serie simplă :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}; \tag{6.79}$$

Scanned with OKEN Scanner

— pentru o serie cu frecvenţe absolute :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{m} f_i}} = \sqrt{\frac{1}{\sum_{i=1}^{m} f_i}\Sigma(x_i - \bar{x})^2 f_i}\ ; \qquad (6.80)$$

— pentru o serie cu frecvenţe relative exprimate în procente :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x})^2 f_i^{*}\%}{100}} = \sqrt{\frac{1}{100}\Sigma(x_i - \bar{x})^2 f_i^{*}\%} \qquad (6.81)$$

— pentru o serie cu frecvenţe relative sub formă de coeficienţi :

$$\sigma = \sqrt{\sum_{i=1}^{m}(x_i - \bar{x})^2 f_i^{*}}. \qquad (6.82)$$

În abaterea medie pătratică însă calculată ca o medie pătratică, prin ridicare la pătrat se dă mai mare importanţă abaterilor mai mari în valoare absolută. Abaterea medie pătratică va fi întotdeauna mai mare decît abaterea medie liniară calculată pentru aceeaşi serie de date.

În literatura de specialitate se apreciază că pentru o serie de distribuţie cu tendinţă clară de normalitate abaterea medie liniară este egală cu 4/5 din valoarea abaterii medii pătratice.

Abaterea medie pătratică este un indicator de bază, care se foloseşte în analiza variaţiei, la estimarea erorilor de selecţie, în calcule de corelaţie. Alături de valoarea medie, abaterea medie pătratică reprezintă un parametru ce defineşte diferitele tipuri de distribuţie teoretice şi serveşte, în special, la interpretarea curbei distribuţiei normale.

Abaterea medie pătratică, ca şi abaterea medie liniară, se exprimă în aceleaşi unităţi ca şi caracteristica a cărei variaţie o caracterizează. Din acest motiv, aceşti indicatori se pot folosi la compararea gradului de variaţie numai pentru seriile care se referă la aceeaşi caracteristică statistică. Pentru compararea gradului de variaţie a două sau mai multe caracteristici statistice se foloseşte coeficientul de variaţie.

Coeficientul de variaţie (v) se calculează ca raport între abaterea medie pătratică şi nivelul mediu al seriei. De obicei se exprimă sub formă de procente. Formula de calcul al coeficientului de variaţie este :

$$v = \frac{\sigma}{\bar{x}} \cdot 100. \qquad (6.83)$$

Dacă se cunoaşte numai abaterea medie liniară, atunci se poate calcula coeficientul de variaţie şi pe baza formulei :

$$v_a = \frac{\bar{d}}{\bar{x}} \cdot 100. \qquad (6.84)$$

Scanned with OKEN Scanner

Coeficientul de variație poate să ia valori între 0 și 100. Cu cît are o valoare mai mică, cu atît seria statistică este mai omogenă și, deci, media este mai reprezentativă. Se apreciază că, în cazul unui coeficient de peste 35—40%, media nu mai este reprezentativă și datele trebuie să fie separate în serii componente, pe grupe, în funcție de variația unei alte caracteristici de grupare. Deci se poate afirma că acest indicator (coeficientul de variație) poate fi folosit ca un test de verificare în aplicarea metodei grupării.

Un alt indicator sintetic al variației, cu o largă utilizare în analiza statistică a fenomenelor, este dispersia.

Dispersia unei caracteristici se notează cu σ^2 și se calculează ca o medie aritmetică simplă sau ponderată a pătratelor abaterilor termenilor față de media lor. Deci se mai poate numi și pătratul mediu al abaterilor termenilor față de media lor. Formulele de calcul sînt:

— pentru o serie simplă:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2; \qquad (6.85)$$

— pentru o serie cu frecvențe absolute:

$$\sigma^2 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{m} f_i} = \frac{1}{\sum_{i=1}^{m} f_i}\sum_{i=1}^{m}(x_i - \bar{x})^2 f_i; \qquad (6.86)$$

— pentru o serie cu frecvențe relative exprimate în procente:

$$\sigma^2 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2 f_i^{*}\%}{100} = \frac{1}{100}\sum_{i=1}^{m}(x_i - \bar{x})^2 f_i^{*}\%; \qquad (6.87)$$

— pentru o serie cu frecvențe relative exprimate sub formă de coeficienți:

$$\sigma^2 = \sum_{i=1}^{m}(x_i - \bar{x})^2 f_i^{*}. \qquad (6.88)$$

Comparînd formulele de calcul ale abaterii medii pătratice și ale dispersiei se observă ușor că abaterea medie pătratică implică calcularea dispersiei, din care se extrage apoi rădăcina pătratică, pentru a ajunge la același grad cu caracteristica a cărei variație se studiază.

Dispersia unui caracteristici reprezintă, în același timp, și momentul centrat de ordinul doi și se poate calcula și fără să fie necesar să se calculeze în prealabil abaterile individuale ale variantelor de la media lor.

Scanned with OKEN Scanner

Dacă în formula dispersiei se dezvoltă binomul $(x_i - \bar{x})^2$ se obține:

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n}.$$

Dacă se ține seama că $\bar{x}$ este o mărime constantă, pătratul ei este tot o constantă, atunci se pot scoate în factor și, scriind membrul doi în fracții separate, se obține:

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n}x_i^2}{n} - \frac{2\bar{x}\sum_{i=1}^{n}x_i}{n} + \frac{n\bar{x}^2}{n},$$

în care:

$\frac{\sum_{i=1}^{n}x_i}{n} = \bar{x}$ și, deci, fracția a doua devine:

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n}x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2.$$

În final se poate scrie că dispersia se mai poate calcula și după formula:

$$\sigma^2 = \frac{\sum_{i=1}^{n}x_i^2}{n} - \bar{x}^2 \quad \text{sau} \quad \sigma^2 = \frac{\sum_{i=1}^{n}x_i^2}{n} - \left(\frac{\sum_{i=1}^{n}x_i}{n}\right)^2. \tag{6.89}$$

Această formulă se poate scrie prin analogie și pentru o serie de frecvențe:

$$\sigma^2 = \frac{\sum_{i=1}^{m}x_i^2 \cdot f_i}{\sum_{i=1}^{m}f_i} - \left(\frac{\sum_{i=1}^{m}x_i f_i}{\sum_{i=1}^{m}f_i}\right)^2. \tag{6.90}$$

De subliniat că, în cazul folosirii unei serii cu intervale de grupare, mărimea dispersiei nu se poate determina ca o valoare exactă. Cu cît intervalele de grupare sînt mai mari, cu atît valoarea dispersiei calculată pe baza centrelor de interval va prezenta o abatere mai mare față de dispersia calculată direct din variantele caracteristicii.

Corecția lui Sheppard se folosește pentru a diminua valoarea dispersiei calculată pentru o serie cu intervale de grupare la care valorile variabilei (x) sînt calculate presupunînd că în interval frecvențele s-ar distribui la

Scanned with OKEN Scanner

distanţe egale şi că media lor ar fi egală cu media aritmetică simplă a celor două valori extreme (limitele intervalului). Sheppard consideră că, în anumite condiţii, dispersia se poate corecta scăzînd 1/12 din pătratul intervalului de grupare. Dispersia astfel corectată (σ^2) va fi egală cu :

$$(\sigma^2)' = \sigma^2 - \frac{h^2}{12}, \qquad (6.91)$$

în care h reprezintă mărimea intervalului de grupare.

El condiţionează aplicarea acestei corecţii numai pentru seriile care prezintă următoarele proprietăţi :

— repartiţia de frecvenţă este continuă ;

— frecvenţele tind către zero în ambele direcţii ;

— numărul cazurilor observate este mai mic de 1 000 unităţi, iar gruparea este mai largă decît cea în care intervalul de variaţie reprezintă a douăsprezecea parte din amplitudinea variaţiei.

Aceste condiţii presupun o repartiţie normală sau uşor asimetrică a frecvenţelor, în care se asigură o compensare a abaterilor şi în interior, nu numai pe total. Deci, nu se recomandă să se aplice pentru repartiţii cu grad mare de asimetrie, pentru care aplicarea corecţiei lui Sheppard nu numai că nu reduce din valoarea erorii, dar influenţează chiar în sensul măririi ei. Rezultă că aplicarea corecţiei lui Sheppard trebuie să se facă cu prudenţă, verificînd în prealabil dacă sînt îndeplinite condiţiile de normalitate şi volum.

Pentru exemplificarea calculării indicatorilor sintetici de variaţie se va folosi următoarea serie de distribuţie de frecvenţe :

Tabelul 6.9

Calculul indicatorilor sintetici de variaţie

Grupe de muncitori după % de îndeplinire a normelor	Nr. muncitorilor (f_i)	Centrul intervalului (x_i)	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
0	1	2	3	4	5	6	7
80— 90	15	85	1 275	−23	−345	529	7 935
90—100	30	95	2 850	−13	−390	169	5 070
100—110	70	105	7 350	− 3	−210	9	630
110—120	60	115	6 900	7	420	49	2 940
120—130	15	125	1 875	17	255	289	4 335
130—140	10	135	1 350	27	270	729	7 290
	200 Σf_i	—	21 600 $\Sigma x_i f_i$	—	−945 +945 1 890 $\Sigma(x_i - \bar{x})f_i$	—	28 200 $\Sigma(x_i - \bar{x})^2 f_i$

Procentul mediu al îndeplinirii normelor :

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i} = \frac{21\,600}{200} = 108\% \text{ muncitor.}$$

Scanned with OKEN Scanner

Abaterea medie liniară :

$$\bar{d} = \frac{\sum_{i=1}^{m} |x_i - \bar{x}| f_i}{\sum_{i=1}^{m} f_i} = \frac{1\,890}{200} = 9{,}45\%.$$

Dispersia :

$$\sigma^2 = \frac{\sum_{i=1}^{m} (x_i - \bar{x})^2 f_i}{\sum_{i=1}^{m} f_i} = \frac{28\,000}{200} = 141.$$

Abaterea medie pătratică :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{m} (x_i - \bar{x})^2 f_i}{\sum_{i=1}^{m} f_i}} = \sqrt{141} = 11{,}87\%.$$

Coeficientul de variaţie :

$$v = \frac{\sigma}{\bar{x}} \cdot 100 = \frac{11{,}87}{108} \cdot 100 = 10{,}99 \simeq 11\%$$

Pe baza rezultatelor obţinute se poate afirma că media este reprezentativă pentru seria din care s-a calculat, deoarece s-a obţinut un coeficient mic de variaţie (11,87%), iar cele mai multe valori (130 din 200) se concentrează în intervalul 100%—120%. Tendinţa de normalitate a distribuţiei se poate constata şi pe baza relaţiei dintre abaterea medie pătratică şi abaterea medie liniară ; 4/5 din abaterea medie pătratică sînt egale cu 9,48 care prezintă o diferenţă minimă (0,03) faţă de abaterea medie liniară (9,45%).

6.3.3. Calculul simplificat al dispersiei şi abaterii medii pătratice. Proprietăţile dispersiei

Calculul dispersiei pe baza abaterilor variantelor înregistrate de la media lor presupune efectuarea unor calcule mai dificile, în special cînd se lucrează cu valori mari ale caracteristicii şi cînd media nu este un cît exact.

Ca şi în cazul mediei se folosesc unele proprietăţi ale dispersiei, care permit o analogie între calculul simplificat al dispersiei şi al mediei.

Scanned with OKEN Scanner

Principalele proprietăţi ale dispersiei, utilizate frecvent în calculele statistice, sînt :

1° Dispersia calculată pentru un şir de valori egale între ele este egală cu zero, deoarece media este egală cu fiecare din variantele înregistrate.

$$\sigma_x^2 = 0; \quad \text{dacă} \quad x_1 = x_2 = \ldots = x_n = \bar{x}. \tag{6.92}$$

2° Dacă fiecare termen al unei serii statistice se modifică într-un sens sau altul cu o mărime constantă (a) şi se calculează dispersia noii serii, atunci ea este egală cu dispersia seriei iniţiale.

Fie seria iniţială :

$$x_1,\ x_2,\ x_3 \ldots x_n$$

pentru care $\bar{x} = \dfrac{\Sigma x_i}{n}$ şi $\sigma_x^2 = \dfrac{\Sigma(x_i - \bar{x})^2}{n}$.

Seria modificată prin scăderea sau adunarea unei constante a :

$$x_1' = x_1 \pm a; \quad x_2' = x_2 \pm a \ldots x_n' = x_n \pm a,$$

pentru care :

$$\bar{x}' = \frac{\Sigma x_i'}{n} = \frac{\Sigma(x_i \pm a)}{n} = \bar{x} \pm a;$$

$$\sigma_{x'}^2 = \frac{\Sigma(x_i' - \bar{x}')^2}{n} = \frac{\Sigma[(x_i \pm a) - (\bar{x} \pm a)]^2}{n} =$$

$$= \frac{(x_i \pm a - \bar{x} \pm a)^2}{n} = \frac{\Sigma(x_i - \bar{x})^2}{n};$$

$$\sigma_{x'}^2 = \sigma_x^2 \text{ dacă } x_i' = x_i \pm a \quad \text{şi } \bar{x}' = \bar{x} \pm a. \tag{6.93}$$

Deci, aplicînd această proprietate, media se modifică, în timp ce dispersia rămîne neschimbată, deoarece nu s-a produs decît o translaţie a valorilor într-un sens sau altul, după cum s-au micşorat sau s-au mărit valorile seriei cu constanta a.

Această proprietate se poate folosi ca un procedeu de calcul simplificat, în special, pentru seriile care au valori mari ale caracteristicii şi, alegînd o nouă origine a variaţiei, să se obţină valori mai mici pentru care este mai uşor de calculat media.

3° Într-o serie de variaţie, dacă se împart sau se înmulţesc toţi termenii cu un coeficient ($K > 1$), dispersia noii serii este de (K^2) ori mai mică sau mai mare decît dispersia seriei iniţiale.

În cazul simplificării cu un coeficient (K) :

— seria iniţială $x_1,\ x_2,\ x_3, \ldots, x_n$,

Scanned with OKEN Scanner

pentru care : $\bar{x} = \frac{\Sigma x_i}{n}$ $\sigma_x^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$;

— seria simplificată :

$$x_1'' = \frac{x_1}{K}; \quad x_2'' = \frac{x_2}{K} \ldots x_n'' = \frac{x_n}{K},$$

pentru care :

$$\bar{x}'' = \frac{\Sigma x''}{n} = \frac{\Sigma \frac{x_i}{K}}{n} = \frac{1}{K} \cdot \frac{\Sigma x_i}{n} = \frac{\bar{x}}{K};$$

$$\sigma_x^2 = \frac{\Sigma(x'' - \bar{x}'')^2}{n} = \frac{\Sigma\left(\frac{x_i}{K} - \frac{\bar{x}}{K}\right)^2}{n} = \frac{\Sigma\left(\frac{x_i - \bar{x}}{K}\right)^2}{n}$$

$$\left(\frac{\Sigma(x_i'' - \bar{x}'')^2}{n}\right)\frac{1}{K^2} = \frac{\sigma_x^2}{K^2}. \tag{6.94}$$

$$\sigma_{x''}^2 = \frac{\sigma_x^2}{K^2}.$$

Deci, aplicînd această proprietate, se modifică în proporţii diferite atît media, cît şi dispersia, adică :

$$\bar{x}'' = \frac{\bar{x}}{K} \quad \text{şi} \quad \sigma_{x''}^2 = \frac{\sigma_x^2}{K^2}. \tag{6.95}$$

Această proprietate se aplică atunci cînd este convenabil să se reducă din volumul calculelor, alegînd o valoare arbitrară, care este divizor comun pentru toţi termenii.

4° Dispersia calculată faţă de o constantă a este mai mare decît dispersia seriei iniţiale cu pătratul diferenţei dintre medie şi constanta a :

$$\sigma_{x-a}^2 = \sigma_x^2 + (\bar{x} - a)^2. \tag{6.96}$$

În cazul folosirii acestei proprietăţi, abaterile se calculează nu faţă de medie, ci faţă de o valoare aleasă arbitrar, notată cu a. Înseamnă că aceste abateri se descompun în abaterea fiecărei variante faţă de medie, plus diferenţa dintre medie şi constanta a :

$$x_i - a = (x_i - \bar{x}) + (\bar{x} - a).$$

Scanned with OKEN Scanner

Deci :

$$\sigma^2_{x-a} = \frac{\Sigma(x_i - a)^2}{n} = \frac{\Sigma[(x_i - \bar{x}) + (\bar{x} - a)]^2}{n}.$$

Dezvoltînd binomul din numărător se obţine :

$$\sigma^2_{x-a} = \frac{\Sigma[(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - a) + (\bar{x} - a)^2]}{n}.$$

Dacă se scriu sumele de la numărător, pe fracţii separate, avînd acelaşi numitor (n) se obţine :

$$\sigma^2_{x-a} = \frac{\Sigma(x_i - \bar{x})^2}{n} + \frac{\Sigma 2(x_i - \bar{x})(\bar{x} - a)}{n} + \frac{\Sigma(\bar{x} - a)^2}{n}.$$

Dacă $\bar{x}$ şi a sînt valori constante, atunci şi diferenţa dintre ele este tot o constantă care se poate da factor comun în fracţia a doua. În cea de-a treia fracţie pătratul diferenţei $(\bar{x} - a)$ este şi el o valoare constantă şi suma se poate înlocui cu numărul unităţilor pentru care se face însumarea. Prima fracţie reprezintă tocmai dispersia seriei iniţiale. Efectuînd aceste operaţii se obţine :

$$\sigma^2_{x-a} = \sigma^2_x + \frac{2(\bar{x} - a)\Sigma(x_i - \bar{x})}{n} + \frac{n(\bar{x} - a)^2}{n}.$$

Pe baza proprietăţilor valorilor medii $\Sigma(x_i - \bar{x}) = 0$ se obţine

$$\sigma^2_{x-a} = \sigma^2_x + (\bar{x} - a)^2.$$

Deci, aplicînd această proprietate, atît media cît şi dispersia se modifică în proporţii diferite :

$$\frac{\Sigma(x_i - a)}{n} = \bar{x} - a \quad \text{şi} \quad \frac{\Sigma(x_i - a)^2}{n} = \sigma^2_x + (\bar{x} - a)^2.$$

De regulă, aceste ultime proprietăţi se combină şi se folosesc în aceleaşi condiţii ca şi la calculul simplificat al mediei. Dacă intervalele sînt egale, atunci se ia a egal cu centrul intervalului cu frecvenţa cea mai mare, iar K se ia egal cu mărimea intervalului de grupare.

Formula de calcul simplificat al dispersiei este :

$$\sigma^2_x = \frac{\sum_{i=1}^{m} \left(\frac{x_i - a}{K}\right)^2 f_i}{\sum_{i=1}^{m} f_i} \cdot K^2 - (\bar{x} - a)^2, \qquad (6.97)$$

Scanned with OKEN Scanner

iar abaterea medie pătratică va fi:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{m}\left(\frac{x_i - a}{K}\right)f_i}{\sum_{i=1}^{m} f_i} \cdot K^2 - (\bar{x} - a)^2}. \qquad (6.98)$$

În exemplul precedent se obține:

Tabelul 6.10

Calculul simplificat al dispersiei

Gr. de muncitori după gradul de îndeplinire a normelor	Nr. muncitorilor f_i	$\frac{x_i - a}{K}$	$\left(\frac{x_i - a}{K}\right)^2$	$\left(\frac{x_i - a}{K}\right)^2 f_i$
0	1	2	3	4
80— 90	15	−2	4	60
90—100	30	−1	1	30
100—110	70	0	0	0
110—120	60	1	1	60
120—130	15	2	4	60
130—140	10	3	9	90
	200 Σf_i	—	—	300 $\Sigma\left(\frac{x_i - a}{k}\right)^2 f_i$

$\bar{x} = 108\%$.

$$\sigma_x^2 = \frac{300}{200} 10^2 - (108 - 105)^2 = 150 - 8 = 141.$$

Deci s-a obținut aceeași dispersie ca și în cazul precedent.

Calculul simplificat al dispersiei se folosește ori de cîte ori este avantajos să se aplice, fie folosind ambele proprietăți, fie numai o singură proprietate.

5° Dispersia calculată pe baza frecvențelor simplificate cu același coeficient este egală cu dispersia seriei inițiale. Această proprietate se bazează pe faptul că dispersia este un pătrat mediu al abaterilor, deci prezintă și ea aceleași proprietăți ca și media aritmetică și care au fost demonstrate în paragraful privitor la proprietățile mediei aritmetice. Această proprietate se aplică atunci cînd dispersia se calculează cu ajutorul frecvențelor relative:

$$\sigma_x^2 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2 \frac{f_i}{\sum_{i=1}^{m} f_i}}{\sum_{i=1}^{m} \frac{f_i}{\sum_{i=1}^{m} f_i}} = \frac{\frac{1}{\sum_{i=1}^{m} f_i}\Sigma(x_i - \bar{x})f_i}{\frac{1}{\sum_{i=1}^{m} f_i}\sum_{i=1}^{m} f_i} = \sigma_x^2 \qquad (6.99)$$

Scanned with OKEN Scanner

6° În cazul cînd colectivitatea este împărţită în grupe distribuţia pe întreaga colectivitate este formată din distribuţiile stabilite pentru fiecare grupă în parte. Dispersia seriei pe întreaga colectivitate este egală cu media dispersiilor componente, plus dispersia dintre mediile parţiale ale tuturor grupelor şi media colectivităţii totale. Această proprietate este folosită frecvent în analiza statistică a variaţiei fenomenelor şi va fi tratată separat în paragraful următor.

6.3.4. Indicatorii variaţiei într-o colectivitate împărţită în grupe. Regula adunării dispersiilor

Cu cît fenomenele sînt mai complexe, cu atît gradul de variaţie al caracteristicilor care le definesc prezintă o amplitudine mai mare a variaţiei. Din această cauză unităţile la care s-a făcut observarea trebuie împărţite în grupe, în funcţie de variaţia factorilor determinanţi.

În cazul în care s-a aplicat în prealabil metoda grupării, atunci se pot calcula atît medii pe grupe, cît şi o medie a colectivităţii totale. În mod corespunzător se vor calcula indicatori de variaţie pentru fiecare grupă şi indicatori de variaţie pe întreaga colectivitate. Indicatorii de variaţie pe întreaga colectivitate se pot calcula fie făcînd abstracţie de faptul că ea este compusă din mai multe grupe, fie luînd în calcul variaţia din interiorul grupelor şi între grupe.

Între indicatorii de variaţie calculaţi la nivelul fiecărei grupe şi cei pe întreaga colectivitate există anumite relaţii, bazate pe regula adunării dispersiilor.

Presupunînd că s-au înregistrat datele pentru o caracteristică x_j şi unităţile au fost împărţite în r grupe, s-au obţinut următoarele distribuţii condiţionate de factorul de grupare (vezi tabelul 6.11).

Tabelul 6.11

Distribuţiile condiţionate ale variabilei x_j în funcţie de un factor de grupare

Grupele colectivităţii	Valorile individuale ale caracteristicii x_1 $x_2 \dots x_j \dots x_m$	Frecvenţele pe grupe	Media pe grupe (media condiţionată)
1	$f_{11} \cdots f_{12} \cdots f_{1j} \cdots f_{1m}$	n_1	$\bar{x}_1$
2	$f_{21} \cdots f_{22} \cdots f_{2j} \cdots f_{2m}$	n_2	$\bar{x}_2$
⋮	⋮	⋮	⋮
i	$f_{i1} \cdots f_{i2} \cdots f_{ij} \cdots f_{im}$	n_i	$\bar{x}_i$
⋮	⋮	⋮	⋮
r	$f_{r1} \cdots f_{r2} \cdots f_{rj} \cdots f_{rm}$	n_r	$\bar{x}_r$
Frecvenţele total	$f_1 \quad f_2 \cdots f_j \cdots f_m$	$\sum_{i=1}^{r} n_i = \sum_{j=1}^{m} f_j = N$	$\bar{x}_0$

Scanned with OKEN Scanner

Deci, frecvenţele pe fiecare grupă se obţin prin însumarea frecvenţelor din interiorul grupelor, adică:

$$\sum_{j=1}^{m} f_{1j} = n_1 ; \ldots \sum_{j=1}^{m} f_{ij} = n_i ; \ldots \sum_{i=1}^{m} f_{rj} = n_r \qquad (6.100)$$

Pentru o astfel de situaţie se obţine o medie generală care sintetizează atît variaţia tuturor valorilor individuale ale colectivităţii totale, cît şi valorile mediilor de grupă (medii condiţionate de factorul de grupare).

După cum rezultă din tabelul 6.11, se pot calcula trei feluri de indicatori care să caracterizeze:

— variaţia valorilor (x_j) în jurul mediei lor de grupă ($x_j - \bar{x}_i$);

— variaţia valorilor mediilor de grupă în jurul mediei colectivităţii totale ($\bar{x}_i - \bar{x}_0$);

— variaţia valorilor (x_j) în jurul mediei colectivităţii totale ($\bar{x}_i - \bar{x}_0$).

La nivelul fiecărei unităţi observate variaţia totală ($x_i - \bar{x}_0$) se poate descompune în variaţia faţă de media de grupă ($x_j - \bar{x}_i$) plus variaţia mediei de grupă de la media colectivităţii totale ($\bar{x}_i - \bar{x}_0$). adică:

$$x_j - \bar{x}_0 = (x_j - \bar{x}_i) + (\bar{x}_i - \bar{x}_0)$$

În sens statistic, variaţia unui ansamblu de fenomene de masă este rezultatul unor factori întîmplători, variabili de la o unitate statistică la alta. Dacă colectivitatea prezintă diferenţieri calitative, atunci înseamnă că şi factorii esenţiali au acţiune variabilă de la o grupă la alta şi, prin urmare, pe lîngă variaţia din interiorul fiecărei grupe există şi variaţia între grupe.

Presupunînd că factorul de grupare este variabil, înseamnă că variaţia totală, în acest caz, este rezultatul combinării acţiunii factorilor întîmplători prezenţi în interiorul fiecărei grupe cu cei esenţiali de la o grupă la alta.

Pentru a măsura, în acest caz, gradul de variaţie provocat de acţiunea combinată a celor două categorii de factori variabili, se foloseşte metoda analizei dispersionale bazată pe descompunerea dispersiei.

Pornind de la ipoteza că media este valoarea tipică, care presupune influenţa tuturor factorilor ca fiind constantă, înseamnă că orice abatere faţă de medie măsoară o anumită mărime a variaţiei provocată de factorii a căror acţiune este prezentă în cazurile înregistrate pentru calculul mediei şi care, potrivit legii numerelor mari, se compensează reciproc.

Media aritmetică a colectivităţii totale, presupunînd toţi factorii constanţi, asigură compensarea tuturor abaterilor valorilor înregistrate faţă de valoarea sa. Dacă media colectivităţii totale presupune toţi factorii esenţiali ca fiind cu acţiune constantă, înseamnă că orice valoare care se abate într-un sens sau altul cuprinde influenţa factorilor întîmplători, cît şi a celor esenţiali variabili şi se măsoară prin abaterea $x_j - \bar{x}_0$.

Factorul de grupare — considerat ca unul din factorii esenţial variabili — influenţează asupra valorii caracteristicii studiate în măsura în care mediile de grupă diferă între ele. Cu cît factorul de grupare are un

Scanned with OKEN Scanner

caracter mai hotărîtor pentru variaţia caracteristicii studiate, cu atît mediile de grupă vor diferi în mai mare măsură de la o grupă la alta. Dacă caracteristica de grupare nu reprezintă însă un factor determinant al caracteristicii studiate, înseamnă că mediile de grupă sînt foarte aproape de valoarea mediei pe total, ceea ce demonstrează că variaţia caracteristicii studiate este produsă de alţi factori, în afara celui ales pentru gruparea datelor. În acest caz, înseamnă că gruparea este numai un mijloc de sistematizare a datelor, după un criteriu organizatoric, ales în funcţie de natura specifică a caracteristicii.

În mod logic înseamnă că variaţia valorilor individuale din fiecare grupă în jurul mediei lor va măsura gradul de influenţă a factorilor variabili prezenţi în interiorul grupei, deoarece, la nivelul grupei, valoarea caracteristicii de grupare este considerată constantă. La rîndul său, variaţia mediilor de grupă faţă de media colectivităţii totale este interpretată ca fiind rezultatul factorului de grupare, deoarece în acest caz factorii variabili din interiorul grupei s-au păstrat la un nivel constant.

Toate aceste variaţii nu se pot calcula numai pentru cîteva cazuri izolate, ci pe întreaga colectivitate, folosind indicatorii de dispersie.

Dispersia totală (σ_0^2) se calculează pe baza tuturor abaterilor individuale stabilite faţă de media colectivităţii totale, după formula:

$$\sigma_0^2 = \frac{\sum_{j=1}^{m} (x_j - \bar{x}_0)^2 f_j}{\sum_{j=1}^{m} f_j}. \qquad (6.101)$$

Media generală, considerînd toţi factorii constanţi, înseamnă că dispersia totală măsoară întreaga variaţie a caracteristicii studiate, produsă de factorii esenţiali şi neesenţiali, înregistraţi şi neînregistraţi.

Cu cît acest indicator are o valoare mai mare, cu atît colectivitatea are un caracter mai eterogen şi este influenţată de un număr mai mare de factori esenţiali şi neesenţiali, constanţi sau variabili, prezenţi de la o grupă la alta, sau în cadrul aceleiaşi grupe.

Dispersia de grupă sau *dispersia parţială* (σ_i^2) se calculează pe baza abaterilor tuturor variantelor dintr-o grupă faţă de media lor, ponderate cu frecvenţele de grupă $\left(\sum_{j=1}^{m} f_{ij} = n_i\right)$.

$$\sigma_i^2 = \frac{\sum_{j=1}^{m} (x_j - \bar{x}_i)^2 f_{ij}}{\sum_{j=1}^{m} f_{ij}} = \frac{\Sigma (x_j - \bar{x}_i)^2 f_{ij}}{n_i} \qquad (6.102)$$

Acest indicator permite măsurarea gradului de influenţă a factorilor variabili prezenţi în interiorul fiecărei grupe. De aceea, acest indicator se mai numeşte şi dispersia din interiorul grupei şi cu cît are o valoare mai mare, cu atît grupa este mai puţin omogenă, apropiindu-se de gradul de variaţie al colectivităţii totale.

Scanned with OKEN Scanner

Pentru a sintetiza această variație într-un singur indicator calculat pe întreaga colectivitate se calculează media dispersiilor parțiale.

Media dispersiilor parțiale ($\overline{\sigma}^2$) se calculează ca o medie aritmetică a dispersiilor parțiale. Dacă grupele sînt egale ca volum, se folosește media aritmetică simplă; dacă volumul diferă de la o grupă la alta se calculează ca o medie aritmetică ponderată.

Media dispersiilor parțiale se poate scrie în două variante:

— ca medie a dispersiilor parțiale:

$$\overline{\sigma}^2 = \frac{\sum_{i=1}^{r} \sigma_i^2 n_i}{\sum_{i=1}^{r} n_i}; \qquad (6.103)$$

— ca sumă a variației din interiorul grupelor raportate la volumul tuturor grupelor:

$$\overline{\sigma}^2 = \frac{\sum_{i=1}^{r} \sum_{j=1}^{m} (x_j - \bar{x}_i)^2 f_{ij}}{\sum_{i=1}^{r} n_i}.^1 \qquad (6.104)$$

Dispersia dintre grupe (δ^2) se calculează pe baza abaterilor mediilor de grupă de la media colectivității totale și măsoară gradul de influență a factorului de grupare asupra variației caracteristicii studiate. Dispersia dintre grupe se calculează după formula:

$$\delta^2 = \frac{\sum_{i=1}^{r} (\bar{x}_i - \bar{x}_0)^2 n_i}{\sum_{i=1}^{r} n_i}. \qquad (6.105)$$

---

1 Această formulă se poate obține dacă la media dispersiilor se scrie explicit formula de calcul, adică:

$$\overline{\sigma}^2 = \frac{\sum_{i=1}^{r} \sigma_i^2 n_i}{\sum_{i=1}^{r} n_i} = \frac{\sum_{i=1}^{r} \sum_{j=1}^{m} (x_j - \bar{x}_i)^2 f_{ij}}{\sum_{i=1}^{r} n_i}, \text{ în care } n_i = \sum_{j=1}^{m} f_{ij},$$

deci:

$$= \frac{\sum_{i=1}^{r} \sum_{j=1}^{m} (x_j - \bar{x}_i) f_{ij}}{\sum_{i=1}^{r} n_i};$$

Între cei trei indicatori există următoarea relaţie : dispersia colectivităţii totale este egală cu media dispersiilor parţiale, plus dispersia dintre grupele colectivităţii, adică :

$$\sigma_0^2 = \overline{\sigma}^2 + \delta^2. \tag{6.106}$$

Această relaţie, cunoscută sub denumirea de regula adunării dispersiilor, se poate determina ca urmare a faptului că cei trei indicatori au acelaşi numitor (suma frecvenţelor condiţionate de factorul de grupare este egală cu suma frecvenţelor variantelor înregistrate independent de variaţia caracteristicii de grupare).

$$\sum_{i=1}^{r} n_i = \sum_{j=1}^{m} f_j. \tag{6.107}$$

Pentru a verifica relaţia dintre cele trei dispersii se va porni de la descompunerea dispersiei totale în funcţie de variaţia mediei condiţionate de la o grupă la alta, adică :

$$x_j - \bar{x}_0 = (x_j - \bar{x}_i) + (\bar{x}_i - \bar{x}_0).$$

Înlocuind în formula dispersiei totale se va obţine :

$$\frac{\sum_{j=1}^{m}(x_j - \bar{x}_0)^2 f_j}{\sum_{j=1}^{m} f_j} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{m}[(x_j - \bar{x}_i) + (\bar{x}_i - \bar{x}_0)]^2 f_{ij}}{\sum_{i=1}^{r}\sum_{j=1}^{m} f_{ij}} =$$

$$= \frac{\sum_{i=1}^{r}\sum_{j=1}^{m}[(x_j - \bar{x}_i)^2 + 2(x_j - \bar{x}_i)(\bar{x}_i - \bar{x}_0) + (\bar{x}_i - \bar{x}_0)^2] f_{ij}}{\sum_{i=1}^{r}\sum_{j=1}^{m} f_{ij}} =$$

$$= \frac{\sum_{i=1}^{r}\sum_{j=1}^{m}(x_j - \bar{x}_i)^2 f_{ij} + \sum_{i=1}^{r}\sum_{j=1}^{m} 2(x_j - \bar{x}_i)(\bar{x}_i - \bar{x}_0) f_{ij} + \sum_{i=1}^{r}\sum_{j=1}^{m}(\bar{x}_i - \bar{x}_0)^2 f_{ij}.}{\sum_{i=1}^{r}\sum_{j=1}^{m} f_{ij}}$$

Dacă se ţine seama că $\sum_{i=1}^{m} f_{ij} = n_i$ şi la nivelul fiecărei grupe $(\bar{x}_i - \bar{x}_0)$ este constant, înseamnă că, scriind pe trei fracţii separate, se obţine :

$$\frac{\sum_{i=1}^{r}\sum_{j=1}^{m}(x_j - \bar{x}_i)^2 f_{ij}}{\sum_{i=1}^{r} n_i} + \frac{2(\bar{x}_i - \bar{x}_0)\sum_{j=1}^{m}(x_j - \bar{x}_i) f_{ij}}{\sum_{i=1}^{r} n_i} + \frac{\sum_{i=1}^{r}(\bar{x}_i - \bar{x}_0) f_i n_i}{\sum_{i=1}^{r} n_i}.$$

Scanned with OKEN Scanner

Conform proprietății mediei aritmetice $\sum_{j=1}^{m}(x_j - \bar{x}_i)f_{ij} = 0$ din fracția a doua, deci întreaga fracție se anulează și în final se obține:

$$\frac{\sum_{j=1}^{m}(x_j - \bar{x}_0)^2 f_j}{\sum_{j=1}^{m} f_j} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{m}(x_j - \bar{x}_i)^2 f_{ij}}{\sum_{i=1}^{r} n_i} + \frac{\sum_{i=1}^{r}(\bar{x}_i - \bar{x}_0)^2 n_i}{\sum_{i=1}^{r} n_i},$$

adică:

$$\sigma^2_{\text{a}} = \overline{\sigma}^2 + \delta^2,$$

în care $\overline{\sigma}^2$ se calculează după varianta a doua.

Media dispersiilor parțiale este:

$$\overline{\sigma}^2 = \sigma_0^2 - \delta^2,$$

iar dispersia dintre grupe este egală cu dispersia totală minus media dispersiilor parțiale:

$$\delta^2 = \sigma_0^2 - \overline{\sigma}^2.$$

Deci este suficient să se cunoască două dintre cele trei dispersii ce intră în relația de bază ca să se determine și a treia.

Aceste relații se folosesc pentru verificarea exactității calculelor și pentru cazul cînd nu se cunosc decît două din cele trei dispersii.

Relațiile dintre abaterea medie pătratică calculată pe total, abaterea medie pătratică din interiorul grupelor, calculată pe baza dispersiilor parțiale și abaterea medie pătratică dintre grupe, se realizează prin intermediul relațiilor dintre dispersiile corespunzătoare, deoarece abaterea medie pe total este egală cu rădăcina pătratică din suma mediei dispersiilor parțiale și a dispersiei dintre grupe, adică:

$$\sigma_0 = \sqrt{\overline{\sigma}^2 + \delta^2}. \qquad (6.108)$$

În seriile statistice formate pe baza frecvențelor relative, indicatorii medii de variație se calculează cu ajutorul formulelor în care frecvența se prezintă sub formă de greutăți specifice exprimate în %, iar suma lor este egală cu 100.

În acest caz, dispersia totală se calculează după formula:

$$\sigma_0^2 = \frac{\sum_{j=1}^{m}(x_j - \bar{x}_0)^2 f_j^{*}\%}{100}, \qquad (6.109)$$

în care: $f_j^{*}\%$ reprezintă frecvențele relative în colectivitatea totală ale variantelor înregistrate independent de factorul de grupare.

Scanned with OKEN Scanner

Dispersia dintre grupe se calculează după formula:

$$\delta^2 = \frac{\sum_{i=1}^{r} (\bar{x}_i - \bar{x}_0)^2 n_i^*\%}{100}, \qquad (6.110)$$

în care: $n_i^*\%$ reprezintă greutatea specifică a fiecărei grupe faţă de total.

Media dispersiilor parţiale

$$\overline{\sigma}^2 = \frac{\sum_{i=1}^{r} \sigma_i^2 n_i^*\%}{100}. \qquad (6.111)$$

În cazul folosirii frecvenţelor relative trebuie să se cunoască atît structura seriilor componente, cît şi structura pe grupe a colectivităţii totale.

După cum s-a mai arătat, frecvenţele relative sînt folosite în cadrul unei singure grupe sau pe întreaga colectivitate, ca şi frecvenţele absolute. Dacă colectivitatea este împărţită în grupe, atunci pentru a calcula media dispersiilor şi dispersia dintre grupe trebuie să se cunoască structura pe grupe a colectivităţii. Dacă grupele sînt egale între ele ca volum, atunci atît media dispersiilor parţiale, cît şi dispersia dintre grupe se calculează ca medie şi, respectiv, dispersie pentru o serie simplă. Dacă volumul diferă de la o grupă la alta, atunci trebuie să se folosească frecvenţele pe grupe.

Indicatorii de variaţie se pot calcula pe baza frecvenţelor relative, cu condiţia să se respecte relaţiile cantitative între frecvenţele absolute şi cele relative, să se ţină seama ca valoarea absolută a unui procent al ponderilor folosite să fie aceeaşi pentru toate elementele componente ale colectivităţii.

6.3.5. Calculul indicatorilor variaţiei unei caracteristici alternative

La calcularea indicatorilor de variaţie ai unei caracteristici alternative trebuie avut în vedere că media ei se calculează ca o greutate specifică a unităţilor care posedă caracteristica.

Pornind de la distribuţia de frecvenţă a caracteristicii alternative, se adaptează formulele generale de calcul ale dispersiei pentru caracteristica alternativă, care au fost prezentate anterior.

Tabelul 6.12

Distribuţia de frecvenţe (absolute şi relative) a caracteristicii alternative

Variantele caracteristicii alternative	Frecvenţele variantelor	
	în cifre absolute (f)	în cifre relative (f^*)
$x_1 = 1$	$f_1 = M$	$f_1^* = p$
$x_2 = 0$	$f_2 = N - M$	$f_2^* = q$
Total	$\Sigma f = N$	$\Sigma f^* = 1$

Scanned with OKEN Scanner

Pentru caracteristica alternativă $\bar{x} = p$.

Folosind frecvenţele relative şi înlocuind pe $(\bar{x})$ cu (p) se obţine dispersia caracteristicii alternative (σ_p^2).

$$\sigma_p^2 = \frac{(1-p)^2 p + (0-p)^2 q}{p+q}.$$

Deoarece $p + q = 1$, atunci $1 - p = q$ şi formula devine:

$$\sigma_p^2 = q^2 p + p^2 q = pq(p+q) = pq = p(1-p).$$

Deci, *dispersia caracteristicii* alternative este egală cu:

$$\sigma_p^2 = pq \quad \text{sau} \quad \sigma_p^2 = p(1-p). \tag{6.112}$$

Abaterea medie pătratică a caracteristicii alternative se va calcula extrăgînd rădăcina pătrată din dispersia:

$$\sigma_p = \sqrt{pq}; \quad \sigma_p = \sqrt{p(1-p)}. \tag{6.113}$$

Frecvenţa relativă a unei caracteristici alternative (p) poate să ajungă pînă la 50% din cazurile observate, respectiv $p = q$, ceea ce înseamnă că dispersia caracteristicii alternative nu poate depăşi 0,25.

În cazul cînd colectivitatea este împărţită în grupe şi se studiază variaţia unei caracteristici alternative, atunci se calculează dispersiile din fiecare grupă, media lor, dispersia dintre grupe şi dispersia colectivităţii totale, după formulele:

Dispersia de grupă pentru o caracteristică alternativă: $(\sigma_{p_i}^2)$:

$$\sigma_{p_i}^2 = p_i q_i \quad \text{sau} \quad \sigma_{p_i}^2 = p_i(1-p_i) \tag{6.114}$$

în care:

p_i reprezintă mediile caracteristicii alternative din fiecare grupă;
q_i — frecvenţele relative ale unităţilor care nu posedă caracteristica în fiecare grupă.

Media dispersiilor parţiale $(\overline{\sigma}_p^2)$ se calculează ca o medie ponderată a dispersiilor grupelor în care a fost împărţită colectivitatea:

$$\overline{\sigma}_p^2 = \frac{\sum_{i=1}^{r} (p_i q_i) N_i}{\sum_{i=1}^{r} N_i}, \tag{6.115}$$

în care:

N_i reprezintă numărul total al unităţilor observate în fiecare grupă.

Scanned with OKEN Scanner

Dispersia dintre grupe pentru o caracteristică alternativă (δ_p^2) se calculează pe baza abaterilor mediilor caracteristicii alternative din fiecare grupă şi media pe întreaga colectivitate:

$$\delta_p^2 = \frac{\sum_{i=1}^{r} (p_i - p)^2 N_i}{\sum_{i=1}^{r} N_i} \tag{6.116}$$

în care:

p reprezintă media caracteristicii generale pe întreaga colectivitate.

Dispersia totală a caracteristicii alternative (σ_p^2) se calculează pe baza celor două greutăţi specifice (q, p) din colectivitatea totală:

$$\sigma_p^2 = pq. \tag{6.117}$$

Regula adunării dispersiilor se păstrează şi în cazul unei caracteristici alternative:

$$\sigma_p^2 = \overline{\sigma}_p^2 + \delta_p^2. \tag{6.118}$$

De unde:

$$\overline{\sigma}_p^2 = \sigma_p^2 - \delta_p^2 \quad \text{şi}$$

$$\delta_p^2 = \sigma_p^2 - \overline{\sigma}_p^2.$$

În mod corespunzător se vor calcula şi abaterile medii pătratice extrăgînd rădăcina pătrată din dispersia generală, dispersia dintre grupe şi media dispersiilor parţiale.

Indicatorii variaţiei unei caracteristici alternative se folosesc pe larg în statistica social-economică în calculul erorilor de selecţie şi la controlul calităţii producţiei, în diferite cercetări cu caracter sociologic. De asemenea, orice caracteristică nealternativă se poate interpreta şi ca o caracteristică alternativă, prin separarea unităţilor în două grupe, sub şi peste o anumită valoare a caracteristicii (x) pentru care se calculează indicatorii de variaţie.

6.3.6. Variaţia intercuartilică şi interdecilică

Pentru a completa analiza seriilor de distribuţie trebuie să se ia în considerare şi forma în care se repartizează frecvenţele în funcţie de valorile individuale care au fost înregistrate. Pentru aceasta este necesar să se folosească, ca valoare tipică, nu numai media aritmetică faţă de care să se studieze variaţia caracteristicii ci şi valorile medii de poziţie: mediana, cuartilele, decilele.

Se ştie că într-o serie perfect simetrică cuantilele — denumirea generalizată a mediilor de poziţie — se distribuie în mod simetric în ambele

Scanned with OKEN Scanner

sensuri față de valoarea tendinței centrale a seriei, calculată ca valoare mediană. Calculînd abaterile dintre valorile mediilor de poziție și valoarea mediană se poate interpreta tendința de distribuție a frecvențelor de apariție ale variantelor caracteristicii.

În astfel de serii, în care s-au calculat valorile cuartilelor, abaterea dintre cuartila inferioară și mediană este egală cu abaterea dintre cuartila superioară și mediană, iar în interiorul lor se găsesc 50% din numărul cazurilor înregistrate.

Ținînd seama de ordinea de creștere a valorilor celor trei cuartile pentru o serie perfect simetrică, această egalitate se poate scrie:

$$Me - Q_1 = Q_3 - Me. \tag{6.119}$$

În acest caz, media aritmetică a celor două cuartile extreme este egală cu valoarea cuartilei a doua, adică cu mediana seriei:

$$\bar{Q} = \frac{Q_1 + Q_3}{2} = Q_2 = Me. \tag{6.120}$$

Dacă cele două relații nu se verifică, adică $Me - Q_1 \neq Q_3 - Me$ și, respectiv, $\bar{Q} \neq Me$, înseamnă că seria prezintă un anumit grad de variație intercuartilică, care trebuie să fie măsurat statistic.

Indicatorii de variație intercuartilică se calculează în mărimi absolute și în mărimi relative.

Abaterea intercuartilică (Q_d) se calculează ca o medie a celor două abateri ale quartilelor extreme față de cuartila centrală:

$$Q_d = \frac{(Me - Q_1) + (Q_3 - Me)}{2} = \frac{Q_3 - Q_1}{2}. \tag{6.121}$$

Înseamnă că abaterea intercuartilică se mai poate numi și *amplitudinea semiintercuartilică*, deoarece se bazează numai pe relația dintre cele două cuartile extreme.

Ca orice indicator absolut și abaterea intercuartilică se exprimă în unitățile de măsură ale caracteristicii studiate și nu poate fi supusă direct comparației statistice a mai multor serii. De aceea se calculează coeficientul de variație intercuartilică.

Coeficientul de variație intercuartilică (V_q) se calculează ca un raport între abaterea intercuartilică și valoarea mediană:

$$V_q = \frac{Q_d}{Me} = \frac{\frac{Q_3 - Q_1}{2}}{Me}. \tag{6.122}$$

Acest coeficient poate să ia numai valori subunitare pozitive și se apreciază că variația intercuartilică este cu atît mai semnificativă, cu cît acest coeficient are o valoare mai mică.

Scanned with OKEN Scanner

Dacă seria prezintă un grad mai mare de asimetrie este necesar să se calculeze și variația interdecilică.

Indicatorii variației interdecilice se bazează pe aceleași considerente, adică într-o serie perfect simetrică distanțele dintre decilele extreme și mediană sînt egale, adică :

$$Me - D_1 = D_9 - Me. \tag{6.123}$$

Abaterea interdecilică este egală cu media aritmetică a abaterilor decilelor extreme față de cuartila centrală a seriei :

$$D_d = \frac{(Me - D_1) + (D_9 - Me)}{2} = \frac{D_9 - D_1}{2}, \tag{6.124}$$

Coeficientul de variație interdecilică se calculează ca raport între abaterea interdecilică și valoarea mediană :

$$V_d = \frac{D_d}{Me} = \frac{\frac{D_9 - D_1}{2}}{Me}. \tag{6.125}$$

De regulă, calculul variației interdecilice se face pentru serii statistice cu un număr mare de grupe și cu tendință evidentă de asimetrie.

Valoarea acestui coeficient se interpretează în același sens ca și aceea a coeficientului de variație intercuartilică.

6.4. ASIMETRIA

În practica statisticii social-economice se pot întîlni serii de distribuție de frecvențe simetrice, ușor asimetrice sau cu tendință pronunțată de asimetrie.

Pentru cazul în care variația este simetrică față de valoarea centrală a caracteristicii, compensarea abaterilor se face nu numai pe ansamblul ei, ci și în interiorul seriei, ca urmare a faptului că frecvențele de apariție ale acestor abateri sînt egale de ambele părți ale valorii centrale. Dacă frecvențele de apariție ale variantelor nu urmează această regularitate, înseamnă că seria prezintă o tendință de asimetrie fie spre valorile mai mari, fie spre valorile mai mici ale caracteristicii.

Seria perfect simetrică corespunde acelei forme de variație statistică în care și influența factorilor întîmplători urmează o anumită regularitate, astfel încît și influențele lor se repartizează în ambele sensuri, iar pe măsură ce crește intensitatea lor scad frecvențele de apariție ale variantelor seriei.

La interpretarea gradului de asimetrie se pornește de la poziția și valorile pe care le au cei trei indicatori ai tendinței centrale : media, mediana

Scanned with OKEN Scanner

și modulul. În special calculul asimetriei se bazează pe relația dintre medie și modul.

Din acest punct de vedere, o serie se poate prezenta în una din cele trei situații :

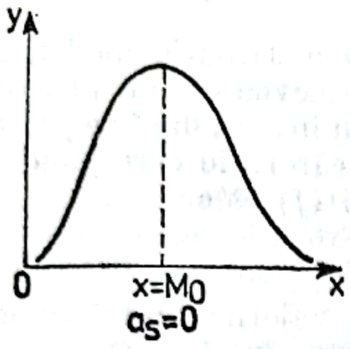


Fig. 6.3. Serie simetrică.

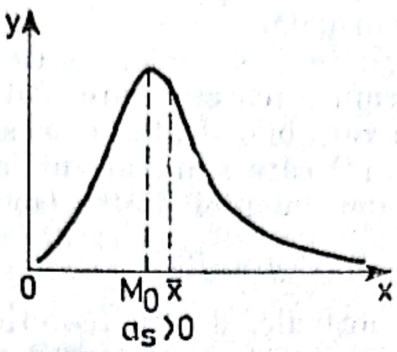


Fig. 6.4. Serie cu asimetrie pozitivă.

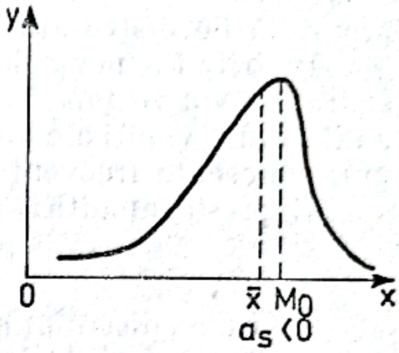


Fig. 6.5. Serie cu asimetrie negativă.

În figura 6.3 se prezintă o serie perfect simetrică în care media este egală cu modulul seriei, ca urmare a faptului că abaterile egale în valoare absolută se compensează reciproc pentru același număr de cazuri observate nu numai pe total.

În figura 6.4 se prezintă o serie asimetrică în care predomină valorile mai mici ale caracteristicii, deci asimetria va fi de sens pozitiv. În acest caz media va fi mai reprezentativă pentru valorile mai mici ale caracteristicii pentru care se vor înregistra abateri mai mici în valoarea absolută, ceea ce înseamnă că vor fi necesare mai multe abateri negative pentru a compensa abaterile mari pozitive, dar mai puțin numeroase.

În figura 6.5 se prezintă o serie cu tendință de asimetrie negativă, deoarece predomină valorile mai mari ale caracteristicii. În acest caz media este mai reprezentativă pentru valorile mari și aici este necesar un număr mai mare de abateri pozitive pentru a compensa un număr mai mic de abateri negative, dar care au o valoare absolută mai mare.

Din cele trei cazuri prezentate reiese că gradul de reprezentativitate al mediei crește pe măsură ce seria se apropie mai mult de distribuția simetrică și are un cîmp mai redus de variație a caracteristicii. De aceea este necesar ca, pe lîngă indicatorii de variație, să se calculeze și indicatorii de asimetrie. La interpretarea indicatorilor de asimetrie trebuie să se țină seama de principalele proprietăți ale distribuției normale.

6.3.1. Interpretarea asimetriei pe baza proprietăților distribuției normale

Distribuția normală (curba lui Laplace-Gauss) face parte din cadrul distribuțiilor în care curba frecvențelor se prezintă sub forma unui clopot în secțiune longitudinală, în care frecvențele de apariție ale unei variabile (x) scad proporțional și simetric în raport cu frecvența maximă, care corespunde valorii centrale. În graficul distribuției normale, ordonata maximă corespunde unui punct pe abscisă, în care valoarea caracteristicii este egală cu media, mediana și modulul seriei.

Legea distribuției normale a fost descoperită de peste 200 de ani, de către De Moivre. La începutul secolului al XIX-lea a fost descoperită

independent de Gauss și Laplace. Proprietățile acestui tip de distribuție sînt deosebit de utile pentru interpretarea statistică a fenomenelor de tip colectiv. Fundamentarea observării selective, aplicarea analizei dispersionale, determinarea intervalelor de încredere, în cadrul cărora se pot plasa diferiți indicatori statistici etc. nu pot fi realizate fără a ține seama de proprietățile distribuției normale.

Distribuția normală poate fi considerată ca o distribuție model, în cadrul căreia se pot determina teoretic care sînt frecvențele de apariție a diferitelor valori ale unei variabile aleatoare (x) sau invers, dacă se presupun cunoscute frecvențele (f) care sînt variantele cărora le corespund.

Expresia analitică a dependenței dintre (x) și (f) este :

$$y = f(x).$$

În cazul distribuției normale, de tip teoretic, valorile caracteristicii (x) pot lua valori de la $-\infty$ la $+\infty$. Pentru a ușura înțelegerea acestui tip de distribuție se consideră, în mod convențional, ca origine a variației caracteristicii (x) valoarea ei medie ($\bar{x} = 0$), iar pe grafic se vor reprezenta abaterile individuale ale caracteristicii (x) față de media seriei ($x - \bar{x}$), ținînd seama de frecvențele lor de apariție. În același timp se vor considera frecvențele de apariție a diferitelor valori ale caracteristicii (x) drept frecvențe relative, care, în cazul distribuției teoretice, se transformă în probabilități de apariție. În cazul distribuției normale teoretice suma probabilităților de apariție a tuturor valorilor caracteristicii (x) este egală cu 1.

În cazul distribuției normale, probabilitățile de apariție a diferitelor abateri ($x - \bar{x}$) considerate ca erori față de medie, corespunzătoare ordonatelor dispuse la un anumit interval față de ordonata maximă, vor fi :

$$y_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - \bar{x})^2}{2\sigma^2}} \qquad (6.126)$$

în care :

y_i reprezintă probabilitatea de apariție a abaterii $(x_i - \bar{x})$;
σ^2 — dispersia caracteristicii x ;
σ — abaterea medie pătratică a caracteristicii x ;
π — 3,14 ;
e — baza logaritmilor naturali = 2,71828 ;
$x_i - \bar{x}$ — abaterea individuală a caracteristicii (x) de la media lor, considerată ca o valoare scontată către care tinde să se concentreze întreaga distribuție de frecvență.

Rezultă că probabilitatea de apariție a abaterii $(x_i - \bar{x})$ egală cu zero va fi :

$$y_0 = \frac{1}{\sigma\sqrt{2\pi}}. \qquad (6.127)$$

Aplicate unor serii empirice, aceste formule de calcul servesc la determinarea frecvențelor relative teoretice de apariție a diferitelor valori ale caracteristicii, considerată ca fiind distribuită normal, dacă se cunoaște media și abaterea medie pătratică.

Scanned with OKEN Scanner

Cînd se cunoaşte volumul colectivităţii (N), în locul frecvenţelor teoretice relative se pot determina frecvenţele teoretice absolute, înlocuind pe 1 (suma frecvenţelor relative) cu N (volumul absolut al întregii colectivităţi). În acest caz, frecvenţa teoretică absolută corespunzătoare unui punct definit pe axa absciselor prin abaterea calculată faţă de medie, va fi :

$$y_t = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{(x_t - \bar{x})^2}{2\sigma^2}}. \qquad (6.128)$$

În cazul distribuţiei normale se demonstrează că între distribuţia dintre diferitele ordonate şi ordonata maximă există o anumită proporţionalitate şi aceste relaţii se găsesc calculate şi tabelate. Ele corespund diferitelor probabilităţi de apariţie a diferitelor valori în raport cu media şi abaterea lor medie pătratică.

Pentru a putea compara mai multe variabile distribuite normal, abaterile ($x_i - \bar{x}$), care sînt măsurate în unităţi concrete de măsură, se exprimă în abateri normale normate. Abaterile normale normate se calculează raportînd abaterile absolute la abaterea medie pătratică :

$$\frac{x_i - \bar{x}}{\sigma}. \qquad (6.129)$$

Aceasta înseamnă că, dacă se cunoaşte media şi dispersia dintr-o cercetare empirică, se pot calcula pentru caracteristica respectivă valorile şi probabilităţile lor de apariţie, măsurate în unităţi de abateri tip şi fără să fie necesară cunoaşterea distribuţiei empirice, cu condiţia ca variabila să urmeze o lege de distribuţie normală a erorilor.

Într-o distribuţie normală, cu probabilităţi de apariţie perfect simetrice într-un sens şi altul, se poate observa pe grafic că punctele de curbură ale poligonului de frecvenţe se găsesc în dreptul ordonatelor corespunzătoare abaterilor măsurate pe baza abaterii medii pătratice.

Astfel se determină valorile abaterilor pe axa absciselor $\bar{x}$; $\bar{x} \pm \sigma$; $\bar{x} \pm 2\sigma$; $\bar{x} \pm 3\sigma$;...cu probabilităţile corespunzătoare.

Calculînd probabilităţile în funcţie de medie şi abaterea medie pătratică, într-o distribuţie normală se poate arăta (vezi egalitatea lui Cebîşev) că în intervalul cuprins între medie şi o abatere medie pătratică ($\bar{x} \pm \sigma$) de ambele sensuri se găsesc 0,6837% din numărul tuturor valorilor posibile, între ($\bar{x} \pm 2\sigma$) se găsesc 0,9345 sau 93,45% cazuri, între ($\bar{x} \pm 3\sigma$) se găsesc 0,9973 sau 99,73% cazuri, iar între ($\bar{x} \pm 4\sigma$) se găsesc 0,9994 sau 99,94% din cazuri. Practic se poate afirma, deci, că în cazul legii de distribuţie normală a erorilor, stabilite ca abateri între valorile caracteristicii şi media lor, un număr foarte mare de cazuri (99 994 din 100 000 posibile) se găsesc cuprinse între valoarea medie şi patru abateri medii pătratice. Acest fapt este deosebit de util pentru interpretarea probabilistică a fenomenelor de masă, care prezintă tendinţă de distribuire normală a frecvenţelor. Pentru astfel de fenomene trebuie avute în vedere proprietăţile distribuţiei normale.

Distribuţia normală este simetrică şi continuă, indică o colectivitate ale cărei caracteristici se manifestă sub forma unor variabile aleatoare cu valori care oscilează în jurul unei valori centrale ce poate fi calculată ca o medie aritmetică din valorile individuale ale caracteristicii. Între

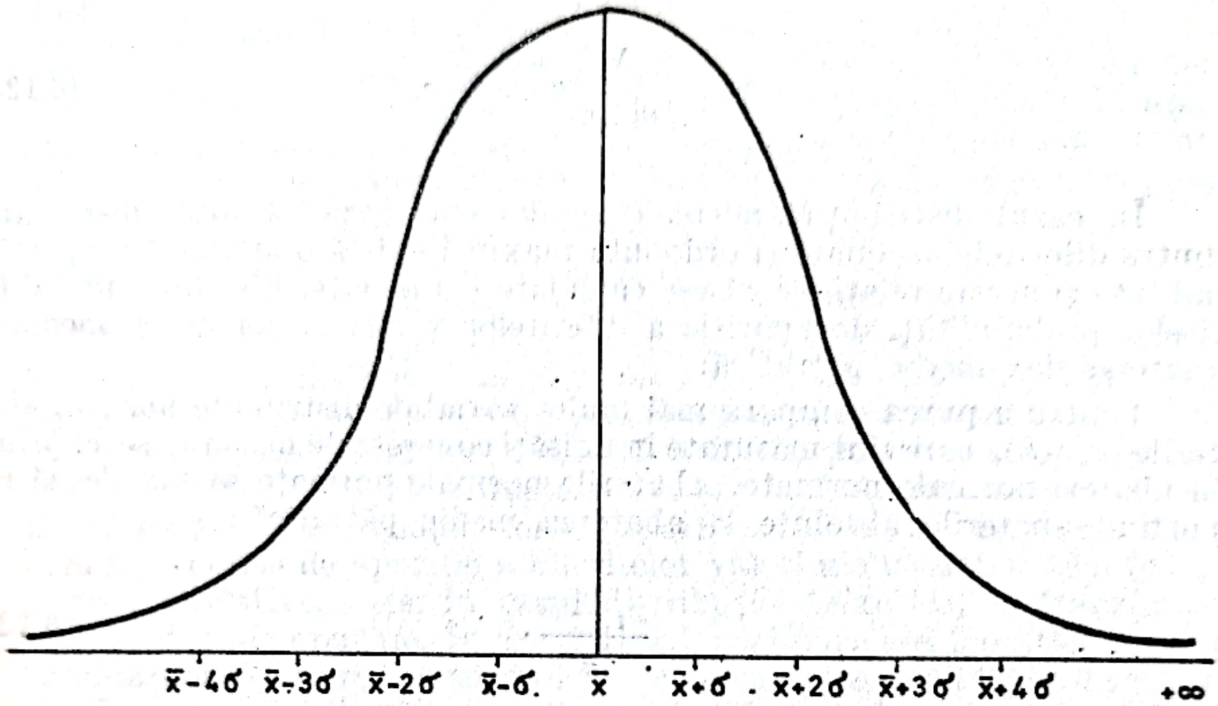


Fig. 6.6. Curba distribuţiei normale.

valorile individuale şi mărimea lor medie există o relaţie de compensare reciprocă a abaterilor într-un sens sau altul, pe baza unor probabilităţi teoretice egale pentru aceleaşi valori absolute ale abaterilor.

În cazul unei distribuţii normale, variabila (x) poate lua orice valoare cuprinsă între $-\infty$ şi $+\infty$, dar cu probabilităţi diferite de apariţie şi anume, ele cresc proporţional şi simetric în ambele sensuri pe măsură ce se apropie de valoarea lor centrală.

Distribuţia probabilităţilor de apariţie a diferitelor valori ale caracteristicii se poate determina şi măsura în unităţi de abateri medii pătratice, deoarece punctele de curbură se găsesc unde ordonatele corespund diferitelor valori ale caracteristicii măsurate prin unităţi de abateri tip. Pe baza acestor relaţii, dacă distribuţia este normală, se pot calcula diferitele probabilităţi corespunzătoare valorilor care trebuie să fie determinate pentru variabila studiată, cunoscîndu-se dinainte media şi abaterea sa medie pătratică, folosind tabelele de calcul ale funcţiei lui Laplace.

Reprezentarea grafică a curbei de distribuţie normală arată că ea depinde de mărimea celor doi parametri care o definesc: media ($\bar{x}$) şi abaterea medie pătratică (σ). Dintre cei doi parametri, influenţa determinantă o are valoarea abaterii medii pătratice, care poate alungi graficul în ambele sensuri, pe cînd valoarea mediei nu produce decît o translaţie pe graficul respectiv.

Cunoaşterea proprietăţilor distribuţiei normale este deosebit de importantă în interpretarea statistică a fenomenelor de masă. Ea se foloseşte, în special, în caracterizarea statistică a seriilor de distribuţie, în ana-

liza dispersională, la verificarea ipotezelor statistice, la efectuarea selecţiilor în vederea caracterizării întregului ansamblu. Pe baza proprietăţilor distribuţiei normale se pot interpreta relaţiile care există între distribuţia empirică obţinută pe baza datelor din observarea statistică şi distribuţia teoretică stabilită pe baza proprietăţilor distribuţiei normale. Această raportare nu se poate face decît dacă variabilele cercetate au tendinţa de distribuţie simetrică a valorilor lor individuale. Pentru a măsura abaterea de la această tendinţă se calculează indicatorii asimetriei.

6.4.2. Indicatorii asimetriei

Indicatorii cu ajutorul cărora se interpretează care este gradul de asimetrie se pot exprima atît în mărimi absolute, cît şi în mărimi relative.

O primă interpretare a gradului şi sensului de asimetrie se bazează pe valorile calculate pentru densitatea de repartiţie a frecvenţelor.

Densitatea de repartiţie a frecvenţelor se calculează ca raport între fiecare frecvenţă şi mărimea intervalului respectiv. Dacă frecvenţa este exprimată în cifre absolute se obţine densitatea absolută de repartiţie; dacă frecvenţa este exprimată în mărimi relative atunci se obţine densitatea relativă a frecvenţelor.

Densitatea absolută de repartiţie a frecvenţelor (d_a) se obţine raportînd frecvenţa absolută (f) la mărimea intervalului (K), adică:

$$d_a = \frac{f_i}{K}. \tag{6.130}$$

Densitatea relativă a frecvenţelor (d_r) se calculează ca raport între frecvenţele relative (f_i^*) şi mărimea intervalului:

$$d_r = \frac{f_i^*}{K}. \tag{6.131}$$

Dacă valorile acestor indicatori arată tendinţa de creştere în valoare către valoarea centrală a caracteristicii, înseamnă că seria este cu tendinţă de normalitate şi că media este reprezentativă pentru cele mai multe valori ale caracteristicii. Necesitatea calculării acestor indicatori apare, în special, pentru seriile cu intervale de grupare mari sau neegale.

Cel mai frecvent însă pentru interpretarea asimetriei se folosesc indicatorii propuşi de Pearson. Cu ajutorul lor se măsoară gradul de oblicitate a seriei.

Coeficientul de asimetrie (C_{as}) se calculează raportînd asimetria absolută ($\bar{x} - Mo$) la abaterea medie pătratică (σ), adică:

$$C_{as} = \frac{\bar{x} - M_0}{\sigma}. \tag{6.132}$$

Acest coeficient de asimetrie poate lua valori între -1 şi $+1$; cu cît este mai mic în valoarea absolută, cu atît asimetria este mai mică.

Scanned with OKEN Scanner

Într-o serie perfect simetrică, coeficientul de asimetrie este egal cu zero, deoarece media coincide în valoare cu modulul seriei. Dacă media este mai mare decît modulul seriei, atunci coeficientul de asimetrie este cuprins între 0 şi 1, deci există o asimetrie pozitivă ; dacă modulul este mai mare decît media, coeficientul de asimetrie este cuprins între —1 şi 0, adică există o asimetrie negativă.

În cazul cînd se cunoaşte mediana seriei, coeficientul de asimetrie (C'_{as}) se poate calcula raportînd abaterea dintre medie şi mediana luată de trei ori, la abaterea medie pătratică :

$$C'_{as} = \frac{3(\bar{x} - M_e)}{\sigma}. \tag{6.133}$$

Această formulă se foloseşte pentru seriile uşor asimetrice în care, pentru un număr mare de cazuri, se verifică relaţia :

$$\text{Modul} = \text{Media} - 3(\text{Media} - \text{Mediana}). \tag{6.134}$$

Acest coeficient poate să ia valori între —3 şi +3 şi va arăta un grad mai mare de simetrie cu cît se va apropia mai mult de zero.

Cei doi coeficienţi de asimetrie se bazează pe relaţia care există între cele trei valori ale tendinţei centrale. Ori, asupra asimetriei influenţează şi celelalte valori medii de poziţie. Pentru a ţine seama şi de influenţa cuartilelor se foloseşte *coeficientul lui Bowley*, calculat după formula :

$$A_s = \frac{q_2 - q_1}{q_2 + q_1}, \tag{6.135}$$

în care :

$$q_2 = Q_3 - M_e;$$

$$q_1 = M_e - Q_1.$$

Rezultă că, în cazul coeficientului lui Bowley, se raportează diferenţa dintre abaterile celor două cuartile extreme de la mediană, la suma acesteia. Bowley apreciază că dacă acest coeficient se apropie de 0,1, seria este moderat asimetrică, iar peste 0,3 seria este pronunţat asimetrică.

Pentru măsurarea gradului de asimetrie şi exces se mai folosesc şi coeficienţii β_1 şi β_2, respectiv γ_1 şi γ_2, propuşi de Pearson, şi care se bazează pe momentele centrate ale seriei [1] :

$$\beta_1 = \frac{\mu'^2_3}{\mu'^3_2}; \quad \beta_2 = \frac{\mu'_4}{\mu'^2_2}, \tag{6.136}$$

---

[1] În statistică se folosesc trei feluri de momente :

a) momentele iniţiale de diferite ordine, atunci cînd punctul de origine al abaterilor este egal cu zero :

$$M_k = \frac{\Sigma x_i^k}{n}; \; M_k = \frac{\Sigma x_i^k f_i}{\Sigma f_i},$$

în care :

M_k reprezintă momentul iniţial de ordinul K. Pentru $K = 1$ $M_1 = \bar{x}_a$;

b) momentele ordinare, cînd punctul de origine al abaterilor este considerat o valoare ar-

în care :

$$\mu_2' = \frac{\Sigma(x - \bar{x})^2}{n} \text{ (momentul centrat de ordinul 2)}; \qquad (6.137)$$

$$\mu_3' = \frac{\Sigma(x - \bar{x})^3}{n} \text{ (momentul centrat de ordinul 3)}; \qquad (6.138)$$

$$\mu_4' = \frac{\Sigma(x - \bar{x})^4}{n} \text{ (momentul centrat de ordinul 4)}; \qquad (6.139)$$

$$\gamma_1 = \pm\sqrt{\beta_1} \quad \gamma_2 = \beta_2 - 3 = \frac{\mu_4' - 3\mu_2'^2}{\mu_2'^2}. \qquad (6.140)$$

Interpretarea statistică a acestor coeficienţi se bazează tot pe referirea la proprietăţile distribuţiei normale. În cazul distribuţiei simetrice, momentele centrate de ordine impare sînt egale cu zero şi pentru seriile asimetrice valoarea lor indică mărimea asimetriei. Deci, cu cît valoarea lui β_1 va fi mai diferită de zero, cu atît gradul de asimetrie este mai mare. Pentru coeficientul β_2 valoarea lui într-o distribuţie normală este egală cu 3 şi interpretarea gradului de aplatizare a curbei se face după cum valoarea calculată pentru o serie empirică este mai mare sau mai mică decît 3. Coeficientul β, ca şi derivatul său γ sînt utilizaţi în cazul seriilor unidimensionale pentru a calcula mărimea excesului.

În general, însă, aceşti coeficienţi (β_1, β_2) necesită calcule destul de laborioase, aşa că se folosesc mai rar în caracterizarea gradului de asimetrie şi exces.

În concluzie, analiza seriilor de repartiţie de frecvenţe se bazează pe un sistem de indicatori care cuprind atît valorile tipice ale seriei, cît şi indicatorii de variaţie şi asimetrie. Pentru compararea seriilor statistice nu sînt suficienţi indicatorii exprimaţi în mărimi absolute, ci trebuie să se calculeze o serie de indicatori relativi sub formă de coeficienţi de variaţie şi asimetrie a fiecărei caracteristici.

Comparînd datele pentru mai multe serii statistice, se poate stabili gradul de omogenitate a unităţilor care au fost cuprinse în calculul medii-

---

bitrară A. Formula de calcul este :

$$\mu_K = \frac{\Sigma(x_i - A)^k}{n}; \qquad \mu_K = \frac{\Sigma(x_i - A)^k f}{\Sigma f};$$

c) momentele centrate cînd punctul de origine al abaterilor este media aritmetică a seriei. Formula de calcul este :

$$\mu_K' = \frac{\Sigma(x_i - \bar{x})^k}{n}; \quad \mu_K' = \frac{\Sigma(x_i - \bar{x})^k f}{\Sigma f}.$$

Deci : $\mu_1' = 0$ şi μ_2' = dispersia.

lor și măsura în care se poate trece, în continuare, la aplicarea metodelor statistice de analiză factorială bazată pe interdependența dintre fenomenele studiate.

Aplicarea diferențiată a metodelor și procedeelor de calcul statistic trebuie să se bazeze pe o analiză teoretică, în care să se țină seama de tendințele obiective de dezvoltare a fenomenelor și relațiile obiective de cauzalitate dintre diferitele caracteristici statistice. Numai pe baza acestei analize se pot interpreta corect indicatorii statistici care au fost calculați și formula concluziile statistice ce au reieșit în urma cercetării.

Indicatorii variației pot fi folosiți la caracterizarea independentă a fenomenelor, la estimarea erorilor de selecție, în analiza corelației statistice și, în general, în toate cazurile cînd se folosesc mărimi medii și trebuie să se interpreteze măsura în care ele sînt reprezentative pentru toți termenii individuali din care au fost calculate.

Scanned with OKEN Scanner

CAPITOLUL VII

CERCETAREA SELECTIVĂ

Folosirea din ce în ce mai frecventă a observării parţiale în cercetarea statistică este legată de unele avantaje pe care le oferă, în raport cu observarea totală. Astfel, utilizarea observării parţiale permite extinderea ariei de cunoaştere statistică şi la fenomenele şi procesele ce nu se pot cerceta printr-o observare totală (determinarea duratei medii de încredere în care se poate consuma un produs alimentar conservat, studiul variaţiei preţurilor produselor pe piaţa ţărănească şi altele). De asemenea, ea se mai poate folosi şi pentru caracterizarea mai amănunţită a unei colectivităţi la care s-a făcut şi observarea totală, sau în cazul organizării unei observări de mare amploare, ca mijloc de control al acesteia. În plus, cercetările parţiale necesită un volum mai mic de cheltuieli de observare şi prelucrare a datelor, în raport cu observarea totală.

Observarea parţială, care răspunde cel mai riguros sarcinilor de cunoaştere statistică a fenomenelor, este observarea selectivă, în care interpretarea rezultatelor selecţiei se face în sens probabilistic şi se pot extinde colectivităţii totale.

7.1. NOŢIUNI INTRODUCTIVE

Cercetarea parţială a cărei scop este ca, pe baza rezultatelor prelucrării datelor obţinute, să se estimeze în anumite condiţii de probabilitate parametrii corespunzători colectivităţii totale poartă denumirea de *cercetare selectivă*. Deci, în cazul utilizării metodei selective, cercetarea statistică cuprinde două părţi :

— prima parte, în care se efectuează observarea şi prelucrarea datelor de selecţie şi

— cea de a doua parte, în care, pe baza datelor de selecţie, indicatorii obţinuţi se extind asupra întregii colectivităţi, în scopul caracterizării ei sub raport statistic.

Din cele arătate reiese că folosirea selecţiei în cercetarea fenomenelor social-economice imprimă cercetării statistice unele particularităţi cu caracter metodologic, care determină utilizarea unor noţiuni specifice.

O primă distincţie trebuie făcută între noţiunea de observare selectivă şi aceea de cercetare selectivă. Observarea selectivă se referă numai la partea de culegere a datelor de la toate unităţile care au intrat în selecţie, pentru toate caracteristicile cuprinse în programul observării, pe cînd cercetarea selectivă cuprinde întregul proces de cercetare, inclusiv partea de extindere şi analiză a rezultatelor.

Cercetarea selectivă, cuprinzînd cele două părţi enunţate mai sus, presupune folosirea unor noţiuni perechi, ca de exemplu, colectivitate de selecţie — colectivitate generală ; medie de selecţie — medie a colecti-

Scanned with OKEN Scanner

vității generale; dispersie de selecție — dispersie a colectivității generale; valoare statistică — parametru și altele.

Colectivitatea generală reprezintă totalitatea unităților simple și complexe din care este format obiectul cercetării statistice concrete. Delimitarea ei în timp și spațiu este necesar să se facă ori de cîte ori se organizează o cercetare statistică, indiferent de forma de observare folosită pentru culegerea datelor individuale.

Colectivitatea de selecție este o parte a colectivității generale la care urmează să se efectueze culegerea datelor în scopul generalizării rezultatelor obținute din prelucrarea acestora asupra întregului ansamblu. Spre deosebire de colectivitatea generală, în care aceleași condiții de timp și spațiu prezintă trăsături esențiale stabile, colectivitatea de selecție poate să varieze de la o selecție la alta, atît ca volum, cît și ca structură. Cu alte cuvinte, indicatorii statistici care definesc colectivitatea de selecție: media, abaterea medie pătratică, coeficientul de variație, coeficienții de asimetrie etc. pot fi considerați și ei de forma unor variabile aleatoare, ce pot reprezenta niveluri diferite de la o selecție la alta, chiar dacă volumul rămîne același.

În teoria și practica statistică, colectivității generale i se mai spune și colectivitate de bază, lot sau populație, iar colectivității de selecție: mostră, eșantion, probă sau, mai simplu, selecție.

Volumul colectivității generale se notează, de regulă, cu N și reprezintă numărul unităților simple care alcătuiesc colectivitatea. Cînd colectivitatea generală este împărțită în grupe, volumul ei $N_i (i = 1,2,\ldots,n)$ se obține ca o sumă a volumelor parțiale ale celor r grupe din care este formată $\left(N = \sum_{i=1}^{r} N_i\right)$.

Volumul colectivității de selecție se notează cu n și reprezintă numărul unităților simple, care formează colectivitatea de la care se culeg datele statistice. Și în cazul colectivității de selecție, dacă sînt mai multe grupe, volumul ei se obține prin însumarea volumului de unități de la fiecare grupă $n_i (i = 1,2,\ldots, r)\left(n = \sum_{i=1}^{r} n_i\right)$.

Nivelul mediu al caracteristicii studiate în colectivitatea generală sintetizează, într-o singură expresie numerică, toate nivelurile individuale ale tuturor unităților $(x_{1,2}\ldots N)$ și se notează cu $\bar{x}_0$, adică:

$$\bar{x}_0 = \frac{\sum_{i=1}^{N} x_i}{N} \tag{7.1}$$

Nivelul mediu al caracteristicii în colectivitatea de selecție sintetizează în valoarea sa numai nivelurile individuale ale unităților care au intrat în eșantion $(x_{1,2}\ldots n)$. Se notează cu $\bar{x}$ și se calculează astfel:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{7.2}$$

În cazul în care se studiază variaţia unei caracteristici alternative, media colectivităţii generale se notează cu p, iar media de selecţie cu w. Frecvenţa unităţilor care posedă caracteristica ($x_1 = 1$) se notează cu M în colectivitatea generală şi cu m în colectivitatea de selecţie. În mod corespunzător, frecvenţa unităţilor care nu posedă caracteristică ($x_2 = 0$) va fi $N-M$, în colectivitatea generală, şi $n-m$, în colectivitatea de selecţie.

Media caracteristicii alternative în colectivitatea generală va fi :

$$p = \frac{M}{N}, \tag{7.3}$$

iar în colectivitatea de selecţie :

$$w = \frac{m}{n}. \tag{7.4}$$

Dispersia colectivităţii generale (σ_0^2) va lua în calcul toate abaterile individuale ale termenilor de la media colectivităţii generale, adică :

$$\sigma_0^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x}_0)^2}{N}, \tag{7.5}$$

iar *dispersia colectivităţii de selecţie* (σ^2) se va calcula numai pentru abaterile individuale ale nivelurilor care au apărut în selecţie faţă de media lor :

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}. \tag{7.6}$$

Pentru caracteristica alternativă, dispersia colectivităţii generale (σ_p^2) va fi :

$$\sigma_p^2 = p \cdot q = p(1-p), \tag{7.7}$$

în care :

p şi q sînt mărimi relative de structură calculate faţă de aceeaşi bază, deci suma lor este egală cu 1.

Analog se calculează dispersia de selecţie (σ_w^2) :

$$\sigma_w^2 = w(1-w). \tag{7.8}$$

În cazul în care colectivitatea generală este împărţită în grupe, intervin în plus notaţiile corespunzătoare grupelor.

O altă pereche de noţiuni frecvent utilizată în selecţie este valoarea statistică şi parametrul. Valoarea statistică a unei caracteristici este aceea

Scanned with OKEN Scanner

calculată pe baza unei distribuţii empirice obţinute prin observare, iar parametrul este valoarea care se estimează pe baza datelor de selecţie. În cadrul cercetării selective, de regulă, media şi dispersia de selecţie se determină ca o valoare statistică, deoarece se calculează pe baza datelor culese prin observarea statistică, iar media şi dispersia colectivităţii generale constituie parametrul care trebuie să fie estimat pe baza datelor selecţiei.

Teoria şi practica statistică au dovedit că între indicatorii obţinuţi prin prelucrarea datelor de selecţie şi cei calculaţi pe baza datelor observării totale pot să apară unele abateri considerate ca erori. În cazul cînd pentru aceeaşi caracteristică se cunosc date din observarea totală şi din cea parţială, aceste erori se calculează ca abateri reale. Cînd se cunosc numai datele de selecţie şi, pe baza lor, se estimează parametrii colectivităţii totale, aceştia apar ca erori probabile. De aceea, pentru fiecare indicator obţinut din prelucrarea datelor de selecţie trebuie calculată şi mărimea erorii sale în raport cu valoarea sa pe întreaga colectivitate. Mărimea erorii de selecţie devine criteriul principal care trebuie să stea la baza organizării unei observări selective, exprimînd gradul de precizie cu care se poate face extinderea rezultatelor selecţiei asupra întregii colectivităţi. Cu cît eroarea de selecţie va fi mai mică, cu atît gradul de precizie al selecţiei va fi mai mare, acesta fiind invers proporţional cu mărimea erorilor ce se produc în această cercetare.

Avantajele selecţiei. Înregistrînd date numai de la o parte a colectivităţii, cercetarea selectivă prezintă unele avantaje faţă de o cercetare statistică bazată pe o observare totală.

Folosind, de regulă, un personal calificat şi ţinînd seama de faptul că numărul unităţilor înregistrate este mult mai mic decît la observarea totală, în observarea selectivă intervin erori de înregistrare mai puţin numeroase decît în cea totală.

Observarea selectivă permite utilizarea unui program de observare mai amănunţit şi datele culese pot da o caracterizare mai cuprinzătoare colectivităţii generale decît în cazul observării totale bazată pe un program de observare mai redus.

Supunînd observării statistice numai o parte din unităţile colectivităţii generale, observarea selectivă se poate utiliza cu succes şi în cazurile în care observarea totală nu se poate aplica (de exemplu, în cazul cercetării calităţii produselor prin analize chimice sau probe fizice, care duc la distrugerea acestor produse).

Efectuîndu-se la un număr mai redus de unităţi, ea este mai operativă şi mai economică, realizîndu-se într-un timp mai scurt, iar culegerea şi prelucrarea datelor se fac cu mai puţine cheltuieli băneşti şi materiale. Astfel, dacă s-ar utiliza observarea totală pentru studierea bugetelor de familie, a preţurilor pe piaţa ţărănească, a resurselor de material lemnos pe esenţe în pădurile ţării etc. ar fi nevoie de un volum de muncă şi de cheltuieli băneşti foarte mari şi care ar fi nejustificate din punctul de vedere al rezultatelor obţinute.

Observarea selectivă este cea mai bună metodă pentru prelucrări preliminare şi verificarea datelor obţinute printr-o observare totală. Ea se utilizează cu succes în prelucrarea preliminară a datelor obţinute printr-o observare totală de mare amploare. De exemplu, datele despre agri-

Scanned with OKEN Scanner

cultură, culese prin recensămîntul din 1948 în ţara noastră, au fost prelucrate mai întîi selectiv şi mai tîrziu total. Diferenţele dintre rezultatele prelucrărilor selective şi cea totală au fost nesemnificative. De asemenea, cu ocazia efectuării unor cercetări ştiinţifice mai amănunţite, în vederea reliefării anumitor aspecte, se pot prelucra selectiv numai unele caracteristici, luate numai de la o parte din unităţile recenzate, care interesează pe cercetător.

Observarea selectivă poate fi folosită pentru testarea unei ipoteze statistice. Astfel, se presupune că s-a emis ipoteza că într-un oraş proporţia celor ce frecventează săptămînal cel puţin un spectacol artistic este de 35 %. Pentru a confirma sau infirma această ipoteză, se organizează una sau mai multe anchete pe baza cărora se calculează valoarea statistică a acestei proporţii, iar abaterile dintre ele sînt luate în considerare la stabilirea mărimii erorii de reprezentativitate.

Domeniile de aplicare a observării selective sînt multiple.

Bazele folosirii în practică a observării selective, denumită şi „anchetă prin sondaj", au fost puse cu aproximativ trei secole mai înainte, fiind legată de unele principii ale statisticii matematice şi ale teoriei probabilităţilor şi, în special, de proprietăţile celor trei repartiţii teoretice (repartiţia binomială introdusă de Bernouli, pe la 1700; repartiţia normală datorată lui Demoivre, dezvoltată mai tîrziu de Laplace şi Gauss; repartiţia Poisson (1857). La aceasta se adaugă contribuţia lui Cebîşev, Leapunov şi alţii.

Avantajele aplicării metodei selecţiei în cercetarea concretă a fenomenelor a făcut ca aria sa de investigaţie să se extindă aproape în toate compartimentele vieţii social-economice. Amploarea şi varietatea domeniilor de aplicare ale selecţiei au dus la apariţia unor institute specializate în domeniul organizării diferitelor anchete selective.

La noi în ţară, utilizarea anchetelor prin sondaj, în sensul modern de observare selectivă, şi-a lărgit aria de aplicare în ultimii 20 de ani şi începînd cu anul 1959 ele îşi găsesc aplicare în toate domeniile vieţii social-economice.

Astfel, în întreprinderile industriale au fost organizate observări selective pentru determinarea gradului de folosinţă a suprafeţelor de producţie; în întreprinderile industriei constructoare de maşini-pentru cunoaşterea gradului de specializare a producţiei, pentru determinarea productivităţii muncii în unităţi naturale, la fotografierea zilei de lucru a muncitorilor şi, într-o măsură din ce în ce mai mare, la controlul calităţii produselor.

În agricultură, observarea selectivă are un domeniu larg de aplicare, începînd cu evaluarea recoltei probabile înainte de recoltare şi pînă la analiza eficienţei economice a introducerii pe scară largă a măsurilor agrotehnice cele mai avansate.

În comerţ, observarea selectivă se foloseşte atît în cadrul pieţei organizate (comerţ socialist) cît şi pentru studierea diferitelor aspecte ale pieţei ţărăneşti. Astfel ea este folosită la studierea cererii de consum a populaţiei, diferenţiată pe cele două forme ale ei, cerere satisfăcută şi cerere nesatisfăcută; pentru confruntarea datelor obţinute prin observarea totală (dări de seamă şi inventarieri); iar pe piaţa ţărănească pentru stabilirea volu-

mului și structurii desfacerilor de mărfuri, precum și pentru cunoașterea variației de prețuri.

Pentru cunoașterea nivelului de trai material și cultural al populației în țara noastră, începînd cu anul 1948 pentru personalul muncitor și din 1952 pentru țărani, Direcția Centrală de Statistică a organizat o observare selectivă denumită metoda bugetelor de familie, care are un caracter de observare curentă, culegînd și prelucrînd date de la un eșantion format inițial din 2 200 bugete de personal muncitor și 3 000 bugete țărănești. În prezent, volumul total al eșantionului este de circa 11 000 bugete de familie, corespunzînd structurii actuale a populației.

Ținînd seama de formele obiective de existență și organizare social-economică a fenomenelor, cu ocazia folosirii unei observări selective se poate face o selecție de unități simple și o selecție de unități complexe. În primul caz se numește selecție individuală, în cel de-al doilea caz selecție de serii, deoarece se consideră că fiecare unitate complexă este formată dintr-o serie de unități simple.

Aplicarea în practică a principiilor selecției la cercetarea concretă a fenomenelor social-economice de masă ridică o serie de probleme legate atît de natura specifică a fenomenelor care trebuie să fie cercetate, cît și unele probleme legate de metodologia de măsurare și comparare a indicatorilor folosiți pentru pregătirea și efectuarea selecției. Cercetarea selectivă, folosindu-se pentru a caracteriza întregul ansamblu, trebuie să fie riguros pregătită, astfel încît informațiile culese să fie cît mai exacte. De aceea este necesar să se întocmească un plan al cercetării selective care să cuprindă :

— delimitarea în timp și spațiu a colectivității totale prin identificarea tuturor cazurilor individuale sub care se manifestă fenomenele respective ;

— verificarea gradului de omogenitate al colectivității generale. Dacă există o observare totală anterioară, aceasta se poate realiza prin analiza indicatorilor variației caracteristicilor esențiale pentru care s-au cules date. Dacă nu există o observare totală, atunci este necesar să se organizeze mai multe sondaje succesive pentru verificarea gradului de stabilitate a mediei și dispersiei, pentru caracteristicile după care se va face eșantionarea ;

— alegerea sau stabilirea bazei de sondaj. Prin bază de sondaj se înțelege orice sistematizare a unităților (liste, hărți), astfel încît să permită alegerea întîmplătoare a unităților ce vor intra în eșantion ;

— alegerea și definirea unității folosite la eșantionare ;

— alegerea tipului și procedeului de selecție ;

— stabilirea periodicității efectuării selecției ;

— stabilirea planului observării, acordîndu-se în special atenție programului observării, care, de regulă, este mai bogat decît cel al unei observări totale similare ;

— stabilirea planului de prelucrare a datelor de selecție din punct de vedere metodologic și organizatoric ;

— alegerea procedeelor de verificare a semnificației indicatorilor de selecție și de extindere a rezultatelor selecției asupra întregului ansamblu.

Scanned with OKEN Scanner

7.2. PROCEDEE DE SELECȚIE

La organizarea în practică a unei observări selective problema esențială o constituie formarea eșantionului.

Alegerea unităților ce vor intra în eșantion se poate face pe baza unei scheme probabilistice sau, cunoscînd dintr-o cercetare anterioară structura și parametrii colectivității generale, se va folosi alegerea intenționată a unităților ce vor intra în selecție. Cu alte cuvinte, în practica eșantionării se pot folosi *selecții aleatoare* sau probabilistice și *selecții dirijate*, sau „subiectiv organizate" sau „alese cu premeditare"

7.2.1. Selecții aleatoare

Selecția aleatoare sau eșantionarea probabilistică este aceea în care, înlăturînd orice intervenție subiectivă, alegerea unităților se face ținînd seama de principiile teoriei probabilităților, care asigură caracterul întîmplător al includerii unităților în eșantion și dă posibilitatea, astfel, să se calculeze limitele erorilor produse față de observarea totală. Trăsătura de bază a selecției aleatoare o constituie faptul că fiecare unitate are o probabilitate egală și diferită de zero, de a participa în selecție, iar probele prin care se obține eșantionul sînt independente între ele.

Selecția aleatoare se poate realiza prin procedeul tragerii la sorți (al loteriei) cu cele două variante, și procedeul selecției bazate pe tabelul cu numere întîmplătoare.

7.2.1.1. PROCEDEUL TRAGERII LA SORȚI

Procedeul tragerii la sorți, presupune amestecarea într-o urnă a tuturor unităților, reprezentate prin bile sau discuri identice ca formă și extragerea la întîmplare a unei bile sau a unui disc pînă ce se completează volumul colectivității de selecție. Acest procedeu se poate folosi în două variante :

— procedeul *selecției repetate* (al bilei revenite) în care, de fiecare dată, bila sau discul se reintroduce în urnă, deci tot timpul cît durează formarea colectivității de selecție toate unitățile au aceeași probabilitate de a intra în eșantion, adică :

$$P = \frac{1}{N}. \tag{7.9}$$

Selecția repetată, datorită faptului că aceeași unitate poate ieși la mai multe extrageri, dă naștere unor erori de selecție mai mari. Această probabilitate este cu atît mai mică, cu cît raportul dintre volumul selecției și cel al colectivității generale este mai mic :

— *procedeul selecției nerepetate* (*al bilei nerevenite*) se bazează pe principiul că o unitate, odată extrasă, nu mai este restituită colectivității generale ; deși șansa fiecărei unități de a intra în colectivitatea de selecție crește

pe măsura efectuării selecţiei, ca urmare a micşorării treptate a volumului colectivităţii generale cu cîte o unitate. La sfîrşitul efectuării selecţiei numărul unităţilor rămase în urnă va fi $N-n$. Probabilităţile de a participa la selecţie vor fi :

(7.10)

$$P_1 = \frac{1}{N}; \quad P_2 = \frac{1}{N-1}; \quad P_3 = \frac{1}{N-2}; \ldots P_n = \frac{1}{N-(n-1)}.$$

În practică, însă, folosirea procedeului tragerii la sorţi este dificil de realizat ţinînd seama, în special, de faptul că în cadrul fenomenelor social-economice se întîlnesc colectivităţi formate din unităţi foarte numeroase, greu de inclus într-o urnă. Date fiind însă avantajele sale, el se foloseşte ori de cîte ori se cercetează colectivităţi mai puţin numeroase şi care prezintă un grad mai mare de omogenitate. De asemenea, el se foloseşte şi în cazul în care colectivitatea este împărţită în grupe şi se cunoaşte volumul selecţiei din fiecare grupă, formarea eşantionului pentru fiecare grupă în parte efectuîndu-se după principiile selecţiei întîmplătoare simple. Selecţia întîmplătoare, repetată sau nerepetată, se mai foloseşte şi în cazul cînd nu se dispune de date cu privire la volumul şi structura unei colectivităţi ; în acest caz se organizează o aşa-zisă „anchetă pilot" în vederea stabilirii unor indicatori pe baza cărora să se poată trece apoi la organizarea unei cercetări propriu-zise.

7.2.1.2. PROCEDEUL TABELULUI CU NUMERE ÎNTÎMPLĂTOARE

Procedeul tabelului cu numere întîmplătoare constă în numerotarea tuturor unităţilor colectivităţii generale într-o ordine oarecare de la 1 la N şi extragerea a n unităţi pe baza unui tabel cu numere întîmplătoare[1].

Tabelul cu numere întîmplătoare este alcătuit cu ajutorul unei maşini de amestecat numere şi aşezarea lor în tabel este întîmplătoare, din orice parte a tabelului s-ar privi.

Eşantionul se formează în felul următor : se stabileşte numărul unităţilor ce trebuie selectate şi numărul total al unităţilor colectivităţii generale. De exemplu, se cercetează o colectivitate generală de 600 unităţi, din care se extrag la întîmplare 20 de unităţi. Unităţile colectivităţii generale vor fi purtătoare ale numerelor de la 1 la 600, deci pentru a forma eşantionul trebuie folosite numere formate cel mult din cîte trei cifre şi care să nu depăşească 600. Se va alege la întîmplare una dintre coloanele tabelului cu numere întîmplătoare cu care se va începe selecţia (vezi tabelul anexă 1). Aceasta poate fi, de exemplu, coloana 3 şi se vor folosi doar ultimele trei cifre ; tot la întîmplare se va alege şi rîndul din coloana respectivă cu care se va începe selecţia. De exemplu, alegînd rîndul al 6-lea din coloana a III-a, la ultimele trei cifre se vor citi : 504, urmează 171 ; 95 ; 916 care nu poate intra în selecţie pentru că depăşeşte numărul 600 ; apoi 132 ş.a.m.d.

---

[1] Tabelele lui Kendall şi Babington-Smith, ale lui Fisher sau Yates.

Scanned with OKEN Scanner

Deci se vor include în eşantion, pe baza tabelului cu numere întîmplătoare, următoarele 20 de unităţi : 504 ; 171 ; 95 ; 132 ; 290 ; 267 ; 81 ; 15 ; 470 ; 505 ; 27 ; 301 ; 409 ; 120 ; 256 ; 387 ; 63 ; 111 ; 491 ; 521.

Se poate observa că în colectivitatea de selecţie sînt incluse la întîmplare unităţi purtînd numere de ordine diferite între 1 şi 600, ceea ce atestă faptul că s-a respectat principiul selecţiei aleatoare, bazat pe folosirea unor probabilităţi egale şi independente.

Cu toate că selecţiile obţinute pe baza acestor procedee respectă cel mai bine principiile selecţiei probabiliste, totuşi nu se pot folosi întotdeauna cu destulă uşurinţă. De aceea, în practica organizării observărilor selective se mai folosesc şi alte procedee derivate din acestea, ca de exemplu, selecţia mecanică.

7.2.1.3. SELECŢIA MECANICĂ

Selecţia mecanică presupune stabilirea unei fracţii de selecţie $K = \frac{N}{n}$[1], care se mai numeşte şi pas de numărare, după care se aleg dintr-o listă unităţile ce vor alcătui eşantionul. Pentru respectarea cît mai riguroasă a principiului introducerii întîmplătoare a unităţilor în eşantion, lista unităţilor trebuie făcută după o caracteristică cu totul întîmplătoare, de pildă ordinea alfabetică sau numerotarea locuinţelor pe străzi etc. După ce s-a stabilit pasul de numărare, din primele K unităţi se alege o unitate la întîmplare, eventual prin tragere la sorţi, la care, adăugînd mereu pasul de numărare, se va forma colectivitatea de selecţie. De exemplu, dacă volumul colectivităţii generale este de 2 000, iar cel al colectivităţii de selecţie de 100, atunci $K = 20$. Prin tragere la sorţi, din primele 20 de unităţi s-a ales unitatea numerotată cu 5, atunci prin selecţie se va include în eşantion următoarele unităţi : 5, 25, 45, 65,...

Dar în cazul selecţiei mecanice probele nu mai sînt independente între ele, ca în cazul selecţiei întîmplătoare simple, fiecare unitate selectată depinzînd de numărul de ordine al celei precedente. Din această cauză selecţia mecanică este considerată, în general, în literatura de specialitate ca fiind cvasialeatoare, în care caracterul întîmplător al selecţiei este asigurat prin alcătuirea întîmplătoare a listelor utilizate în selecţie. Selecţia mecanică este frecvent utilizată în agricultură şi în unele cercetări cu caracter sociologic, pe baza unor liste care există în mod frecvent pentru evidenţa curentă. De exemplu, se pot folosi ca bază de selecţie, pentru procedeul mecanic, listele existente la consiliile populare privitor la autoimpunere sau listele electorale.

7.2.2. Selecţii dirijate

Spre deosebire de selecţiile probabilistice, în eşantioanele dirijate alegerea unităţilor se face în mod frecvent conştient de către cei ce culeg datele cu privire la colectivitatea cercetată. Un exemplu de astfel de selecţie poate fi considerat eşantionul format pe baza aplicării *procedeului*

[1] Pasul de numărare K se ia totdeauna cu o valoare întreagă obţinută prin rotunjire.

cotelor (quotasampling). Acest procedeu de selecţie se poate aplica atunci cînd se cunoaşte bine structura colectivităţii generale şi din fiecare grupă tipică se stabileşte o cotă de eşantion. Alegerea unităţilor în cadrul cotelor este făcută pe teren cu condiţia ca unităţile alese să se încadreze grupei tipice. Alcătuirea colectivităţii de selecţie în cadrul fiecărei cote de eşantion depinde de voinţa recenzorului, şi deci, nu asigură fiecărei unităţi a colectivităţii generale şanse egale de a pătrunde în selecţie. Eşantionarea pe cote, nebazîndu-se pe probabilităţi egale ale unităţilor de a pătrunde în selecţie, nu poate fi considerată ca selecţie probabilistică. Cu toate acestea, pe ansamblu, selecţia pe cote poate fi considerată ca fiind reprezentativă, dacă gruparea prealabilă efectuată în colectivitatea generală a dus la formarea unor grupe cu un grad mare de omogenitate. Dacă în cadrul grupei, gradul de variaţie este mai mic, atunci diferenţa între eşantionarea probabilistică şi cea dirijată este relativ mică.

Datorită faptului că această cercetare este mai ieftină şi mai uşor de realizat din punct de vedere organizatoric, ea este destul de frecvent utilizată în prospectarea pieţei şi în sondajele de opinie, unde interpretarea probabilistă are un cîmp mai mare de manifestare. Acest procedeu de selecţie a fost utilizat şi în ţara noastră, de şcoala sociologică de la Bucureşti, în cercetările concrete efectuate în mediul rural, în perioada dintre cele două războaie mondiale.

În practica eşantionării se pot întîlni cazuri în care, deşi au fost respectate principiile de bază ale selecţiei, totuşi eşantioanele formate au un grad mic de reprezentativitate.

În astfel de împrejurări se pot pune în mod legitim cîteva întrebări.

Este necesar să se efectueze o nouă selecţie, respingîndu-se eşantionul care nu îndeplineşte condiţia de reprezentativitate, păstrînd aceeaşi caracteristică de eşantionare, acelaşi procedeu de selecţie şi acelaşi volum al selecţiei?

Este necesar să se schimbe caracteristica de eşantionare sau procedeul de selecţie pentru a asigura o selecţie reprezentativă?

Trebuie să se continuie selecţia, mărind volumul eşantionului în scopul includerii în eşantion şi a altor unităţi, care să corecteze gradul de reprezentativitate?

Desigur, că cel mai simplu ar fi să se respingă eşantionul ca fiind nereprezentativ şi să se treacă la efectuarea unei noi selecţii, fie folosind aceeaşi caracteristică de eşantionare, fie schimbînd-o. Dar teoria şi practica arată că este mult mai convenabilă alcătuirea de eşantioane repetate prin care poate fi măsurată caracteristica aleatoare a procedurii.

În ceea ce priveşte creşterea volumului eşantionului, în general, se apreciază că aceasta nu poate fi considerată ca principala cale care ar putea duce la creşterea gradului de reprezentativitate a selecţiei. Se poate demonstra uşor că, după un anumit volum al eşantionului, eroarea începe să scadă mult prea lent faţă de creşterea volumului selecţiei.

În practica eşantionării, în cazul obţinerii unor colectivităţi de selecţie nereprezentative, este indicat să fie verificate, în primul rînd, principiile care au stat la baza selecţiei şi apoi în ce măsură ele corespund naturii specifice a fenomenelor ce formează obiectul cercetării şi a posibilităţii de obţinere a informaţiei statistice. În general, în asemenea cazuri se con-

Scanned with OKEN Scanner

sideră că dau rezultatele cele mai bune selecţiile obţinute din colectivităţi împărţite în grupe tipice, formate pe baza unor caracteristici esenţiale de grupare.

7.3. ERORILE CERCETĂRII SELECTIVE

Erorile cercetării selective pot fi de înregistrare, comune oricărui fel de observare statistică, şi de reprezentativitate, care sînt specifice cercetării selective.

Erorile de înregistrare care pot interveni în cadrul unei observări selective, datorită volumului mai mic, sînt mai puţin numeroase şi însemnate decît în cadrul unei observări totale şi pot fi înlăturate prin efectuarea controlului cantitativ şi calitativ obişnuit.

În cercetarea selectivă este important să fie stabilită eroarea de reprezentativitate, care apare ca diferenţă sau abatere între indicatorii derivaţi ce caracterizează colectivitatea de selecţie (mărimi relative, medii, indicatori de variaţie şi corelaţie etc.) şi cei corespunzători calculaţi pentru întreaga colectivitate. Dintre aceste abateri cea mai frecvent estimată este abaterea dintre media de selecţie şi media colectivităţii generale ($\bar{x}-\bar{x}_0$).

7.3.1. Tipuri de erori

După natura lor, erorile de reprezentativitate întîlnite în practica statistică sînt şi ele de două feluri: sistematice şi întîmplătoare. Erorile de reprezentativitate sistematice provin din nerespectarea strictă a principiului de bază al selecţiei întîmplătoare, care cere ca unităţile să fie incluse în eşantion fără nici un fel de preferinţă subiectivă din partea celui care efectuează selecţia.

De exemplu, dacă cercetînd selectiv nivelul productivităţii muncii, la formarea eşantionului se preferă numai unităţile care au nivelul caracteristicii mai ridicat, media calculată din datele de selecţie va fi mult mai mare decît media generală, dînd naştere unei erori sistematice de reprezentativitate în sens pozitiv. De asemenea, dacă în cazul cercetării selective a timpului pierdut în medie de un muncitor, în cadrul unui schimb, se aleg în mod deliberat muncitorii care sînt mai conştiincioşi, atunci timpul mediu, stabilit pe baza datelor de selecţie va fi mult mai mic decît timpul mediu ce se pierde la întreaga colectivitate de muncitori studiată.

Uneori erorile de reprezentativitate pot să apară din comoditatea cercetătorului. De exemplu, în cazul cercetării selective a recoltei probabile este necesar ca toate tipurile de parcele să fie reprezentate în eşantion. Dacă cel ce ia probele, ocoleşte în mod sistematic părţile mai greu accesibile, chiar dacă acestea au căzut în eşantion, recolta medie calculată va avea o abatere mare în plus faţă de recolta medie ce se va stabili la recoltarea efectivă.

Tot în această categorie de erori se poate include și *nonrăspunsul*, adică pentru unele unități care au intrat în selecție nu s-au obținut datele individuale în timpul programat și, deci, nu vor putea fi cuprinse în calculul indicatorilor de selecție.

Este ușor de înțeles că erorile sistematice de reprezentativitate influențează negativ veridicitatea datelor de selecție și că ele pot fi înlăturate dacă se respectă întru totul principiile stabilite de teoria selecției.

Spre deosebire de acestea, erorile întîmplătoare de reprezentativitate nu se pot evita chiar dacă se respectă cu rigurozitate principiile de bază ale teoriei selecției, ca urmare a faptului că în selecție nu se poate reproduce absolut identic colectivitatea generală. Aceasta înseamnă că, în general, pe baza datelor de selecție nu se poate face o estimare absolut corectă a mediei colectivității generale, că efectuînd mai multe selecții de același volum pot să apară valori diferite ale mediei de selecție, care se vor abate într-o măsură mai mare sau mai mică de la media colectivității generale.

Observarea totală, cuprinzînd toate unitățile colectivității generale, ține seama de întreaga influență exercitată de cauzele esențiale și neesențiale, de sensul și gradul de intensitate a acestei influențe, concretizate în apariția unor niveluri diferite de dezvoltare a acestor unități. Înregistrîndu-se nivelurile de la toate unitățile și frecvențele lor de apariție se poate stabili seria de distribuție a variabilei studiate, pe baza căreia se vor calcula indicatorii derivați care o caracterizează. Observarea selectivă, cuprinzînd numai o parte din unitățile colectivității generale, nu poate înregistra absolut toate nivelurile individuale ale caracteristicii cu frecvențele corespunzătoare și, deci, nu poate reproduce *identic* seria de distribuție a variabilei din colectivitatea generală. Repetînd selecția, pentru același volum al ei, fiecare eșantion va fi caracterizat printr-o anumită serie de distribuție, care se va deosebi într-o măsură mai mică sau mai mare de aceea a colectivității generale. Cum media oscilează în jurul valorii cu ponderea cea mai mare, media fiecărui eșantion poate să fie mai mare sau mai mică decît media colectivității generale în funcție de ponderile cu care s-a calculat, dînd în mod corespunzător naștere, ca sens și mărime, unor erori întîmplătoare de reprezentativitate.

Pentru ca erorile de estimare a mediei colectivității totale să fie minime, eșantionul supus cercetării statistice trebuie să îndeplinească *condiția de reprezentativitate*.

Prin *reprezentativitatea colectivității* de selecție se înțelege ca partea supusă observării statistice să reprezinte aceleași trăsături esențiale, reproducînd într-un număr mai mic de unități aceeași structură ca și a întregii colectivități care se studiază. În cazul studierii unor colectivități care prezintă tipuri calitative diferite, deci colectivitatea generală este împărțită în grupe tipice, componența calitativă a părții supuse observării trebuie să corespundă structurii pe grupe a colectivității de bază. Cum în practică nu se pot reproduce structuri identice, se consideră ca realizare a condiției de reprezentativitate a selecției, selecția la care greutățile specifice ale fiecărei grupe nu diferă cu mai mult de $\pm 5\%$, în raport cu structura colectivității de bază.

De reținut faptul că erorile de reprezentativitate se pot calcula ca erori efective numai în cazul în care valorile individuale ale unei caracte-

ristici statistice au fost înregistrate şi în eşantion şi în colectivitatea generală. În cazul în care caracteristicile studiate figurează numai în programul mai larg al observării selective, ele trebuie să fie estimate prin metode de calcul statistic.

Rezultă de aici, că după modul de calcul, erorile de reprezentativitate pot fi : *erori efective* şi *erori probabile*.

7.3.1.1. EROAREA EFECTIVĂ DE REPREZENTATIVITATE

Erorile efective de reprezentativitate se pot determina pentru variabila la care s-a făcut eşantionarea şi reprezintă un criteriu de verificare a reprezentativităţii eşantionului.

De exemplu, pentru a verifica gradul de reprezentativitate a unei observări selective privind cererea de consum a unei grupe de populaţie se efectuează două selecţii succesive.

Este ştiut că cererea de consum depinde, în primul rînd, de mărimea veniturilor obţinute de grupa respectivă de populaţie. Deci, în acest caz se va folosi drept caracteristică de eşantionare, după care se va alcătui colectivitatea de selecţie, mărimea veniturilor, în raport de variaţia căreia se va verifica gradul de reprezentativitate a selecţiei.

Considerînd că volumul total al unităţilor din grupa respectivă de populaţie este de 20 000, că volumul celor două selecţii succesive este de 10%, pe baza centralizării datelor pe grupe de venituri s-au obţinut următoarele serii empirice de distribuţie de frecvenţe :

Tabelul 7.1

Distribuţia muncitorilor după mărimea veniturilor în două eşantioane succesive

(date convenţionale)

Grupe de muncitori după mărimea veniturilor	Distribuţia numărului de muncitori					
	colectivitate generală		în prima selecţie		în a doua selecţie	
	cifre absolute	cifre relative	cifre absolute	cifre relative	cifre absolute	cifre relative
0	1	2	3	4	5	6
sub 2 000	1 000	5,0	60	3,0	124	6,2
2 001—2 200	2 400	12,0	164	8,2	260	13,0
2 201—2 400	5 600	28,0	440	22,0	500	25,0
2 401—2 600	7 400	37,0	840	42,0	760	38,0
2 601—2 800	2 200	11,0	280	14,0	240	12,0
peste 2 800	1 400	7,0	216	10,8	116	5,8
Total	20 000	100,0	2 000	100,0	2 000	100,0

Comparînd cele două distribuţii de selecţie cu distribuţia de frecvenţe a colectivităţii generale se pot constata unele abateri care le imprimă acestora tendinţe diferite. În prima selecţie, valorile mai mari sînt mai bine reprezentate, pe cînd în cea de-a doua selecţie ponderile mai mari au valorile mai mici ale caracteristicii.

Scanned with OKEN Scanner

Pentru a putea face o apreciere asupra modului în care colectivitatea de selecţie reprezintă sau nu colectivitatea generală, se vor calcula abaterile de structură a celor două colectivităţi de selecţie de la structura colectivităţii de selecţie teoretică. Aceste abateri se vor nota cu d_f şi sînt înscrise în tabelul 7.2, coloanele 4 şi 6.

În literatura de specialitate, structurii colectivităţii teoretice de selecţie i se mai spune şi structură programată, deoarece structura ei este identică cu cea a colectivităţii generale, ceea ce face ca eroarea ei de selecţie să fie egală cu zero. Selecţia teoretică (programată) se obţine prin aplicarea frecvenţelor relative ale colectivităţii generale, la volumul colectivităţii de selecţie (vezi tabelul 7.2, coloana 2).

Tabelul 7.2

Verificarea reprezentativităţii eşantionului

Grupe de muncitori după mărimea veniturilor	Selecţia teoretică		Selecţia I		Selecţia II	
	Frecvenţe absolute	Frecvenţe relative	Frecvenţe absolute	Abateri faţă de cele teoretice (d_f)	Frecvenţe absolute	Abateri faţă de cele teoretice (d_f)
0	1	2	3	4	5	6
sub 2 000	100	5,0	60	−40	124	+24
2 001—2 200	240	12,0	164	−76	260	+20
2 201—2 400	560	28,0	440	−120	500	−60
2 401—2 600	740	37,0	840	+100	760	+20
2 601—2 800	220	11,0	280	+60	240	+20
peste 2 800	140	7,0	216	+76	116	−24
Total	2 000	100,0	2 000	$\Sigma(d_f)=0$ $\Sigma\|d_f\|=472$	2 000	$\Sigma(d_f)=0$ $\Sigma\|d_f\|=168$

Rezultă, deci, că pe ansamblu abaterile de structură, la acelaşi volum de selecţie, se compensează reciproc, deoarece $\Sigma d_f = 0$. Pentru a putea măsura gradul de abatere a structurii selecţiei efective de la structura programată se va face raportul dintre suma abaterilor de structură pozitive şi negative, luate sub formă de modul, şi volumul selecţiei (n), adică:

$$K_f = \frac{\Sigma|d_f|}{n} \cdot 100, \qquad (9.11)$$

în care:

K_f reprezintă coeficientul de realizare a structurii programate a colectivităţii de selecţie.

Calculînd pentru exemplul considerat:

$$K_{f.I} = \frac{472}{2\,000} \cdot 100 = 23{,}6\%;$$

$$K_{f.II} = \frac{168}{2\,000} \cdot 100 = 8{,}4\%.$$

Scanned with OKEN Scanner

Aceasta înseamnă că, la prima selecţie, aproape 1/4 din unităţile selectate reprezintă abateri de la structura programată şi care vor provoca erori într-un sens sau altul, în timp ce la cea de-a doua selecţie, coeficientul de abatere este sub 10%. Este de aşteptat ca indicatorii obţinuţi din prelucrarea datelor de selecţie de la cea de-a doua colectivitate să estimeze mai bine parametrii respectivi din colectivitatea generală. Este suficient să se verifice acest lucru prin calcularea mediilor de selecţie şi a mediei colectivităţii generale:

$$\bar{x}_0 = 2\,416 \text{ lei}; \quad \bar{x}_{\text{I}} = 2\,476{,}40 \text{ lei}; \quad \bar{x}_{\text{II}} = 2\,408 \text{ lei}.$$

Eroarea efectivă de reprezentativitate ($d_{\bar{x}}$):

$$d_{\bar{x}} = \bar{x} - \bar{x}_0 \begin{cases} \text{selecţia I} \rightarrow d_{\bar{x}} = 2\,476{,}40 - 2\,416 = 60{,}40 \text{ lei} \\ \text{selecţia II} \rightarrow d_{\bar{x}} = 2\,408 - 2\,416 = -8 \text{ lei}. \end{cases}$$

Pentru a stabili gradul de reprezentativitate a eşantionului pe baza erorii medii de selecţie se raportează această eroare la valoarea mediei colectivităţii totale:

$$d_{\bar{x}} = \frac{\bar{x} - \bar{x}_0}{\bar{x}_0} \begin{cases} \text{selecţia I} \rightarrow \dfrac{60{,}40}{2\,416} \cdot 100 = +2{,}5\% \\ \text{selecţia II} \rightarrow \dfrac{-8}{2\,416} \cdot 100 = -0{,}33\% \end{cases} \tag{9.13}$$

Deci, eşantionul al doilea este mult mai reprezentativ decît primul, deşi amîndouă se încadrează în intervalul admis ($\pm 5\%$).

De cele mai multe ori în cercetarea concretă nu se dispune de date cu privire la întreaga colectivitate, pentru a putea verifica astfel în ce măsură media de selecţie este sau nu reprezentativă. În cazurile în care media pe întreaga colectivitate nu este cunoscută, este necesar să se efectueze mai multe selecţii succesive spre a verifica gradul de stabilitate a mediei şi dispersiei pentru variabila după care se face eşantionarea. În acest caz eroarea de reprezentativitate se va calcula:

$$d_{\bar{\bar{x}}} = \bar{x} - \bar{\bar{x}},$$

în care:

$\bar{\bar{x}}$ reprezintă media mediilor selecţiilor efectuate pentru verificarea stabilităţii ei.

Practica a dovedit că realizarea unei selecţii cu un grad mai mare de reprezentativitate nu se poate face decît în măsura în care selecţia este alcătuită pe baza unei scheme probabiliste, pentru care se pot calcula anticipat limitele erorilor produse cu ocazia aplicării ei şi pentru caracteristicile care nu sînt cuprinse în observarea totală.

Pentru a răspunde acestor cerinţe este necesar ca la alcătuirea colectivităţii de selecţie să se ţină seama de următoarele condiţii:

— probele prin care se formează eşantionul trebuie să fie independente între ele, ca urmare a aplicării stricte a principiului folosirii unor

probabilităţi egale şi diferite de zero la formarea eşantionului pentru toate unităţile colectivităţii generale;

— volumul eşantionului trebuie să fie suficient de mare, astfel încît în el să fie prezentă acţiunea tuturor factorilor care determină gradul şi forma de variaţie a nivelurilor individuale ale caracteristicilor studiate în cadrul colectivităţii generale. Numai în acest fel nivelurile individuale ale unităţilor selectate intră în cîmpul de acţiune a legii numerelor mari, în care abaterile într-un sens sau altul se compensează reciproc.

Reiese, deci, că respectînd aceste condiţii, concluziile formulate pe baza prelucrării datelor de selecţie trebuie considerate ca *enunţuri probabile*, obţinute ca urmare a aplicării metodei selecţiei aleatoare şi pentru care trebuie calculate limitele intervalului lor de încredere.

7.3.1.2. EROAREA MEDIE PROBABILĂ DE REPREZENTATIVITATE

În practica selecţiei nu se pot calcula erori de reprezentativitate efective pentru toate caracteristicile înscrise în programul observării selective; pentru caracteristica la care nu se cunoaşte media colectivităţii generale eroarea de reprezentativitate se calculează pe baza relaţiilor care există între indicatorii de selecţie şi cei ai colectivităţii generale, cu condiţia ca selecţia să fi fost făcută pe baza unei scheme probabilistice.

În teoria selecţiei se demonstrează că, dacă volumul colectivităţii de selecţie este suficient de mare, distribuţia unităţilor în cadrul ei va urma într-o măsură suficient de mare legea distribuţiei din colectivitatea generală, iar media de selecţie, ca expresie sintetică a nivelurilor individuale ale tuturor unităţilor cercetate, va fi cît mai aproape de media colectivităţii generale.

În practică, însă, pentru acelaşi volum de selecţie se pot obţine mai multe eşantioane, în funcţie de volumul colectivităţii totale şi pentru care se pot calcula valori diferite ale mediei de selecţie. În acest proces de formare a mediilor de selecţie, fiecare medie de selecţie, la rîndul ei, poate să apară o singură dată sau de mai multe ori. Se poate spune, deci, că şi media de selecţie este tot o variabilă aleatoare căreia i se poate stabili seria de distribuţie de frecvenţă.

Pentru a putea urmări modul de formare a distribuţiei de eşantionare a mediei se va lua un exemplu, în care volumul colectivităţii generale este de patru unităţi, din care se vor forma toate eşantioanele posibile din cîte două unităţi. Deci $N = 4$, iar $n = 2$. Fiecare unitate va fi notată prin literele A, B, C, D cu următoarele mărimi ale caracteristicii: $A = 12$; $B = 14$; $C = 16$; $D = 18$.

Media colectivităţii generale:

$$\bar{x}_0 = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{12 + 14 + 16 + 18}{4} = 15.$$

Scanned with OKEN Scanner

Dispersia colectivității generale:

$$\sigma_0^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x}_0)^2}{N} = \frac{(12-15)^2+(14-15)^2+(16-15)^2+(18-15)^2}{4} = 5.$$

Dacă se efectuează toate selecțiile posibile, folosind procedeul bilei revenite (selecția repetată) se vor obține următoarele eșantioane, de cîte două unități:

$A+A$	$B+A$	$C+A$	$D+A$
$A+B$	$B+B$	$C+B$	$D+B$
$A+C$	$B+C$	$C+C$	$D+C$
$A+D$	$B+D$	$C+D$	$D+D$

Aceasta arată că se pot forma patru grupe de combinații de cîte două unități. Generalizînd, înseamnă că în cazul selecției repetate se poate efectua un număr de eșantioane egal cu N^n, respectiv în exemplul luat $4^2 = 16$.

În cazul selecției nerepetate (procedeul bilei nerevenite) combinațiile sînt mai puține datorită faptului că aceeași unitate nu poate participa decît într-un singur eșantion. Formula de calcul al numărului de eșantioane posibile pentru selecția nerepetată este dată de formula de calcul al combinărilor, aplicată la selecție:

$$C_N^n = \frac{N!}{n!(N-n)!}.$$

Verificînd în exemplul prezentat, vor apărea următoarele combinații posibile:

$A+B$	$B+C$
$A+C$	$B+D$
$A+D$	$C+D$

adică:

$$C_4^2 = \frac{4!}{2!(4-2)!} = 6.$$

Efectuîndu-se toate aceste selecții posibile (folosind ambele procedee) pentru fiecare selecție se va obține o anumită valoare a mediei cu care se va estima mărimea mediei pe întreaga colectivitate. La rîndul ei fiecare medie va depinde de nivelurile individuale care au intrat în eșantionul respectiv, dînd naștere unei dispersii corespunzătoare (tabelul 7.3).

Scanned with OKEN Scanner

Mediile, erorile şi dispersiile de selecţie folosind procedeul bilei revenite

Denumirea unităţii	Nivelurile caracteristicii în eşantion x_i	Media de selecţie $\bar{x}$	Eroarea de selecţie $\bar{x}-\bar{x}_0$	Dispersia de selecţie σ_i^2
0	1	2	3	4
A+A	12+12	12	−3	0
|A+B|	12+14	13	−2	1
|A+C|	12+16	14	−1	4
|A+D|	12+18	15	0	9
B+A	14+12	13	−2	1
B+B	14+14	14	−1	0
|B+C|	14+16	15	0	1
|B+D|	14+18	16	+1	4
C+A	16+12	14	−1	4
C+B	16+14	15	0	1
C+B	16+16	16	+1	0
|C+D|	16+18	17	+2	1
D+A	18+12	15	0	9
D+B	18+14	16	+1	4
D+C	18+16	17	+2	1
D+D	18+18	18	+3	0

Notă: selecţiile puse în chenar sînt cele ce se pot obţine aplicînd procedeul bilei nerevenite.

Se poate observa că, efectuînd toate combinaţiile posibile, s-au obţinut valori diferite ale mediei şi dispersiei de selecţie, unele apărînd o singură dată, altele de mai multe ori. Centralizînd aceste valori se pot stabili distribuţii de frecvenţe pentru mediile de selecţie, pentru erorile de reprezentativitate pe care ele le produc, ca şi pentru dispersiile corespunzătoare acestor eşantioane.

Dacă se consideră media de selecţie în sens probabilistic, adică fiecare valoare a sa ca eveniment favorabil faţă de toate celelalte evenimente egal posibile, repetînd de un număr de ori selecţia, frecvenţa de apariţie a unei anumite valori a mediei, luată ca frecvenţă relativă, are tendinţa de a coincide cu probabilitatea de producere a erorii sale faţă de valoarea mediei colectivităţii generale. În cazul selecţiei aleatoare se poate demonstra că pentru o anumită fracţie de selecţie $\left(\frac{n}{N}\right)$ fiecărei valori a mediei de selecţie îi corespund probabilităţi de apariţie bine determinate, în funcţie de mărimea absolută a abaterii sale faţă de media colectivităţii generale.

Folosind datele din exemplul prezentat anterior se vor folosi seriile de distribuţia de frecvenţe a mediilor de selecţie, respectiv a erorilor de reprezentativitate, în funcţie de probabilităţile lor de apariţie (tabelul 7.4).

Pentru a uşura interpretarea acestor serii de distribuţie de frecvenţe, ele se vor reprezenta grafic comparîndu-le cu curba distribuţiei normale a abaterilor. Deoarece seria de distribuţie a mediilor de selecţie coincide ca formă cu seria de distribuţie a erorilor de reprezentativitate, pe grafic se vor trece numai valorile erorilor.

Scanned with OKEN Scanner

Tabelul 7.4

Distribuţia de eşantionare a mediilor şi erorilor de selecţie

Nr. crt.	Mediile de selecţie posibile $\bar{x}$	Frecvenţa mediilor de selecţie (f_i)	Eroarea de selecţie ($\bar{x}-\bar{x}_0$)	Probabilitatea de apariţie $P_i = \frac{f_i}{\Sigma f_i}$
0	1	2	3	4
1	12	1	−3	1/16 = 0,06
2	13	2	−2	2/16 = 0,125
3	14	3	−1	3/16 = 0,190
4	15	4	0	4/16 = 0,250
5	16	3	+1	3/16 = 0,190
6	17	2	+2	2/16 = 0,125
7	18	1	+3	1/16 = 0,06
—	—	$\Sigma f_i = 16$	$\Sigma(\bar{x} - \bar{x}_0) = 0$	$\Sigma P_i = 1$

Analizînd aceste serii de distribuţie (vezi tabelul 7.4 şi fig. 7.1) se poate constata că distribuţia mediilor de selecţie şi, respectiv, a erorilor pe care ele le produc se prezintă sub forma unei distribuţii simetrice, în care cea mai mare probabilitate de apariţie o are media de selecţie, care coincide în valoare cu media colectivităţii generale şi pentru care eroarea de reprezentativitate este egală cu zero. Faţă de această valoare centrală, celelalte valori ale mediei de selecţie se distribuie simetric de ambele părţi, cu probabilităţi bine determinate, egale pentru aceeaşi abatere absolută

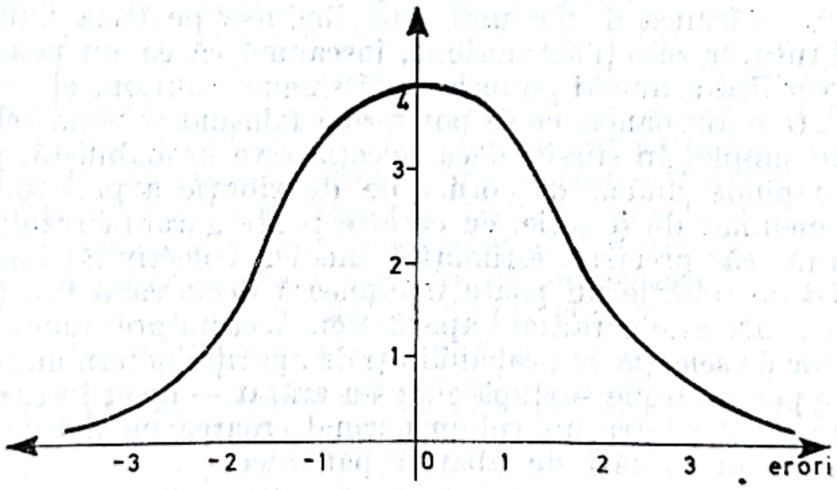


Fig. 7.1. Distribuţia erorilor de selecţie.

într-un sens şi altul. Dacă selecţia este suficient de mare (peste 40 de unităţi), atunci probabilităţile de apariţie a mediilor de selecţie descresc proporţional şi simetric, pe măsură ce se apropie de probabilităţile corespunzătoare celor două valori ale mediei de selecţie. Deci, se poate spune că probabilităţile descresc pe măsură ce cresc erorile de reprezentativitate. Dacă mediile de selecţie se distribuie după legea normală, înseamnă că erorile întîmplătoare de reprezentativitate se pot măsura în abateri nor-

male normate $\left(\frac{\bar{x} - \bar{x}_0}{\sigma_{\bar{x}}}\right)$, iar pe baza proprietăţilor acestei distribuţii se poate afirma că în cazul selecţiei probabilistice 67% dintre erori se găsesc în intervalul cuprins între $\bar{x} \pm \sigma_0$; 95,45% dintre erori sînt cuprinse între $\bar{x} + 2\sigma_0$; 97,7% în intervalul $\bar{x} \pm 3\sigma_0$; iar 99,9% între $\bar{x} \pm 4\sigma_0$. Deci, dacă se cunoaşte legea de distribuţie a mediilor de selecţie, se poate determina cu anticipaţie, cu un anumit grad de siguranţă, pe baza abaterii standard a acestora, care este mărimea erorii de reprezentativitate la care se poate aştepta.

La determinarea mărimii erorilor de reprezentativitate este necesar să se facă deosebirea dintre *exactitate*, *distorsiune* şi *precizie*.

Se spune că o selecţie are un grad mare de exactitate în măsura în care media ei are o abatere minimă faţă de parametrul colectivităţii generale ce urmează să fie estimat pe baza datelor de selecţie. Rezultă, deci, că *gradul de exactitate* a mediei de selecţie se măsoară ca diferenţă între media de selecţie ($\bar{x}$) şi media colectivităţii generale ($\bar{x}_0$), adică $\bar{x} - \bar{x}_0$. Dacă nu se cunoaşte media colectivităţii generale, această abatere nu poate fi considerată ca o abatere reală, ci ca o abatere probabilă ce poate fi interpretată numai în măsura în care formarea colectivităţii de selecţie s-a făcut pe principiile selecţiei probabiliste şi eroarea mediei de selecţie urmează o lege de distribuţie de probabilităţi. În acelaşi timp, dacă selecţia s-a făcut acordînd aceeaşi şansă de a participa la formarea eşantionului fiecărei unităţi a colectivităţii generale, media mediilor de selecţie coincide în valoare cu media colectivităţii generale, deci selecţia este nedistorsionată. *Mărimea distorsiunii* se măsoară tocmai ca diferenţă între media mediilor de selecţie ($\bar{\bar{x}}$) şi media colectivităţii generale ($\bar{x}_0$), respectiv: $\bar{\bar{x}} - \bar{x}_0$. Mărimea distorsiunii, calculîndu-se pe baza unui indicator sintetic al tuturor selecţiilor posibile, înseamnă că ea nu poate fi determinată şi verificată numai pe baza unui singur eşantion, ci luînd în considerare toate eşantioanele ce se pot forma folosind schema selecţiilor întîmplătoare simple. În sfîrşit, dacă selecţia este probabilistă, *precizia estimaţiilor* depinde numai de forma de distribuţie a probabilităţilor de apariţie a mediilor de selecţie, cu care se poate garanta rezultatul selecţiei. Rezultă că precizia estimaţiei mediei colectivităţii generale pe baza mediei de selecţie nu poate fi înţeleasă decît ca o funcţie de probabilitate cu care este garantată apariţia ei. Aceasta presupune o distribuţie a mediilor de selecţie cu probabilităţi de apariţie determinate, distribuţie ce nu se poate obţine — după cum s-a arătat — decît în cazul selecţiei probabiliste, cînd pentru un volum normal eroarea mediei de selecţie se poate măsura în unităţi de abateri pătratice.

Trebuie reţinut, deci, că gradele de exactitate, distorsiune şi de precizie nu se pot verifica numai pe baza unui singur eşantion, ci considerînd selecţia respectivă ca făcînd parte dintr-un proces întreg de formare a tuturor eşantioanelor posibile, în care mediile de selecţie au probabilităţi diferite de apariţie, în funcţie de mărimea erorii de reprezentativitate la care sînt asociate. Aceasta înseamnă că mărimea erorii de reprezentativitate trebuie considerată în funcţie de întreaga distribuţie a mediilor de selecţie şi a abaterii acestora de la media colectivităţii generale. Cum valoarea sa este variabilă de la un eşantion la altul, este necesar ca toate aceste

erori să se sintetizeze într-o singură expresie numerică calculîndu-se astfel valoarea ei medie. Cum mediile de selecţie şi, respectiv, erorile produse de ele tind să se distribuie normal, înseamnă că, efectuînd toate selecţiile posibile, erorile într-un sens şi altul tind să se compenseze reciproc, deci, dacă se ţine seama şi de sensul erorilor, atunci media lor este egală cu zero sau tinde către zero. Ca şi în cazul abaterii medii liniare, pentru a evita aceasta la calculul erorii medii de reprezentativitate se foloseşte media pătratică. Deoarece eroarea de reprezentativitate se determină ca o abatere între fiecare medie de selecţie şi media colectivităţii generale ($\bar{x} - \bar{x}_0$), înseamnă că dacă media mediilor de selecţie coincide cu media colectivităţii generale, atunci *eroarea medie pătratică* poate fi considerată ca o abatere medie pătratică a tuturor mediilor de selecţie de la media lor.

Pentru a deosebi abaterea medie pătratică calculată din abaterile individuale de la media lor (σ) de abaterea medie pătratică a mediilor de selecţie de la media acestora, ultima se va nota cu (μ) şi, pe baza notaţiilor anterioare, formula de calcul va fi:

$$\mu = \sqrt{\frac{\sum_{i=1}^{K} (\bar{\bar{x}} - \bar{x}_0)^2 f_i}{\sum_{i=1}^{K} f_i}}, \qquad (7.14)$$

în care:

K reprezintă numărul mediilor de selecţie posibile;

f — frecvenţa de apariţie a fiecărei medii de selecţie.

Fără îndoială această formulă de calcul exprimă mai bine gradul de exactitate al estimaţiei mediei colectivităţii generale pe baza datelor de selecţie. Dificultatea constă în faptul că, în practică, mărimea distorsiunii este foarte greu de determinat. De aceea, cercetătorul, atunci cînd organizează o observare selectivă, trebuie să ia toate măsurile de evitare a distorsiunii, pentru ca datele culese să intre în cîmpul de acţiune al legii numerelor mari, asigurîndu-se astfel compensarea erorilor. Aceasta înseamnă că toate calculele de determinare cu anticipaţie a erorii de selecţie nu se pot face decît considerînd colectivitatea de selecţie ca fiind nedistorsionată, în care media mediilor de selecţie este egală cu media colectivităţii generale, iar erorile care se produc se compensează reciproc.

Dacă se verifică în continuare aceste principii generale pe baza exemplului considerat, media mediilor de selecţie va fi:

$$\bar{\bar{x}} = \frac{(12 \cdot 1) + (13 \cdot 2) + (14 \cdot 3) + (15 \cdot 4) + (16 \cdot 3) + (17 \cdot 2) + (18 \cdot 1)}{16} = 15.$$

Valoarea medie a mediilor de selecţie ($\bar{\bar{x}}$) fiind egală cu media colectivităţii generale ($\bar{x}_0$) se poate folosi formula de calcul al erorii medii pătratice pentru selecţia nedistorsionată. Erorile de selecţie fiind deja calculate (vezi tabelul 7.5, col. 3), calculul erorii de reprezentativitate va fi:

$$\mu = \sqrt{\frac{(-3)^2 \cdot 1 + (-2)^2 \cdot 2 + (-1)^2 \cdot 3 + (0)^2 \cdot 4 + (1)^2 \cdot 3 + (2)^2 \cdot 2 + (3)^2 \cdot 1}{16}} =$$

$$= \sqrt{2{,}5} \simeq \pm 1{,}58.$$

Deci, pentru cele 16 eşantioane posibile, folosind selecţia întîmplătoare repetată, eroarea de reprezentativitate care poate să apară în medie este de 1,58 unităţi într-un sens sau altul. De remarcat este faptul că eroarea medie de reprezentativitate se exprimă în aceleaşi unităţi de măsură ca şi nivelurile individuale ale caracteristicii cercetate.

Interpretînd conţinutul formulei de calcul al erorii medii de reprezentativitate reiese că ar trebui să se cunoască toate mediile de selecţie posibile, frecvenţele corespunzătoare lor şi media colectivităţii generale pentru a vedea dacă selecţia este distorsionată sau nu. Dar, în practică, metoda selectivă se aplică tocmai pentru a înlocui o observare totală, deci, de regulă, se cunoaşte media de selecţie şi, eventual, o altă medie dintr-o cercetare totală sau parţială anterioară. De aceea, în practică, pentru determinarea erorii medii de reprezentativitate se folosesc anumite relaţii care există între dispersia colectivităţii generale (σ_0^2), dispersia mediilor de selecţie de la media colectivităţii generale (μ^2) şi volumul selecţiei (n).

În teoria selecţiei se demonstrează că, în cazul *selecţiei întîmplătoare repetate*, dispersia colectivităţii generale este egală cu dispersia mediilor de selecţie de la media colectivităţii generale înmulţită cu volumul selecţiei :

$$\sigma_0^2 = \mu^2 \cdot n, \tag{7.15}$$

de unde :

$$\mu^2 = \frac{\sigma_0}{n} \text{ ; iar } \mu = \sqrt{\frac{\sigma_0^2}{n}} = \frac{\sigma_0}{\sqrt{n}} \tag{7.16}$$

Pe baza acestei relaţii se poate afirma că eroarea medie de reprezentativitate este direct proporţională cu mărimea variaţiei, sintetizată în abaterea medie pătratică a colectivităţii generale, şi invers proporţională cu radical din volumul selecţiei. Cum în aceleaşi condiţii de timp şi spaţiu abaterea medie pătratică a colectivităţii generale este o valoare constantă, rezultă că mărimea erorii medii de selecţie poate fi influenţată prin modificarea volumului eşantionului.

În ceea ce priveşte primul factor, trebuie arătat că nu totdeauna se cunoaşte mărimea dispersiei colectivităţii generale (σ_0^2) nedispunîndu-se de date dintr-o observare totală. Se apreciază că dacă, în general, condiţiile esenţiale au rămas aceleaşi şi se dispune de date dintr-o observare totală anterioară, atunci se poate folosi în calculul erorii medii de reprezentativitate această dispersie. Dacă din diferite motive nu există date pentru calcularea dispersiei generale, atunci se pot folosi date din mai multe selecţii experimentale de circa 100 unităţi. În cazul unui eşantion se poate formula ipoteza că dispersia de selecţie este aproximativ egală cu dispersia colectivităţii generale şi atunci eroarea medie pătratică de selecţie va fi :

$$\mu = \sqrt{\frac{\sigma^2}{n-1}},$$

în care :

σ^2 reprezintă dispersia de selecţie ;

$n-1$ — volumul selecţiei corectat cu un grad de libertate al dispersiei de selecţie, în raport cu dispersia colectivităţii totale.

Scanned with OKEN Scanner

În exemplul considerat — avînd calculate dispersia colectivității generale și dispersia mediilor de selecție — relația se verifică astfel:

$$\sigma_0^2 = n \cdot \mu^2$$

$$5 = 2{,}5 \times 2.$$

În concluzie, se poate aprecia că dacă selecția este aleatoare și nedistorsionată, eroarea medie de reprezentativitate se poate calcula pe baza acestei relații, adică:

$$\mu = \sqrt{\frac{\sigma_0^2}{n}} \simeq \sqrt{\frac{\sigma^2}{n-1}}. \qquad (7.17)$$

Dacă cercetătorul vrea să mărească sau să micșoreze eroarea medie, atunci trebuie să acționeze asupra modificării volumului eșantionului; respectiv, va trebui, deci, să împartă sau să înmulțească pe n din formula erorii medii de reprezentativitate cu K^2, adică:

$$\mu K = \sqrt{\frac{\sigma_0^2}{\frac{1}{K^2} n}} \text{ pentru mărire;} \qquad (7.18)$$

$$\frac{\mu}{K} = \sqrt{\frac{\sigma_0^2}{K^2 n}} \text{ pentru micșorare.} \qquad (7.19)$$

În cazul *selecției nerepetate*, numărul de eșantioane este mai mic ca urmare a faptului că aceeași unitate nu poate participa decît la o singură selecție. Din această cauză relația dintre volumul selecției, dispersia mediilor de selecție și dispersia colectivității generale se transformă într-o inegalitate de forma:

$$n\mu^2 \text{ (nerepetat)} < \sigma_0^2. \qquad (7.20)$$

În exemplul luat, eșantioanele obținute pe baza aplicării procedeului selecției întîmplătoare nerepetate vor fi ca în tabelul 7.5.

Tabelul 7.5

Mediile și erorile de selecție folosind procedeul bilei nerevenite

	Media de selecție $\bar{x}$	Eroarea de reprezentativitate $\bar{x}-\bar{x}_0$
0	1	2
$A+B$	13	−2
$A+C$	14	−1
$A+D$	15	0
$B+C$	15	0
$B+D$	16	+1
$C+D$	17	+2

Scanned with OKEN Scanner

$$\mu^2 = \frac{(-2)^2 \cdot 1 + (-1)^2 \cdot 1 + (0)^2 \cdot 2 + (1)^2 \cdot 1 + (2)^2 \cdot 1}{6} = 1{,}67$$

$$n\mu^2 \text{ nerep.} < \sigma_0^2;\ 2 \times 1{,}67 < 5;\ 3{,}34 < 5.$$

În cazul *selecției întîmplătoare nerepetate* această relație apare ca urmare a faptului că probabilitățile de a participa la selecție cresc de la $\frac{1}{N}$ la prima probă, la $\frac{1}{N-(n-1)}$ la ultima extragere, astfel încît volumul rămas este de $N-1$ la selecția repetată, și $N-n$ la selecția nerepetată. Pornind de aici, înseamnă că, în acest caz, dispersia mediilor de selecție (μ^2) trebuie corectată înmulțind cu raportul $\frac{N-n}{N-1}$, și formula erorii medii de reprezentativitate, în cazul selecției întîmplătoare, devine:

$$\mu_{nerep} = \sqrt{\frac{\sigma_0^2}{n}\left(\frac{N-n}{N-1}\right)}. \qquad (7.21)$$

Coeficientul $\sqrt{\frac{N-n}{N-1}}$ se numește coeficient de corecție a erorii medii de reprezentativitate în cazul aplicării procedeului bilei nerevenite. În practică, cînd se organizează o cercetare selectivă după procedeul selecției nerepetate și volumul colectivității de selecție este suficient de mare, nu se mai scade unitatea la numitor și coeficientul devine $\sqrt{1-\frac{n}{N}}$, în care $\frac{n}{N}$ reprezintă tocmai proporția de selecție.

În cazul cînd nu se cunoaște volumul colectivității generale, sau proporția de selecție este foarte mare, la calculul erorii medii de reprezentativitate, indiferent de schema utilizată (repetată sau nerepetată), se folosește formula de calcul de la selecția repetată (adică se renunță la coeficientul de corecție al erorilor).

În cazul cînd se cercetează variația unei caracteristici alternative, relațiile arătate rămîn valabile, folosind în calcul dispersia acestei caracteristici, iar eroarea medie pătratică se va nota cu μ_w pentru a o deosebi de μ_x, care este eroarea medie pătratică pentru o caracteristică nealternativă, respectiv:

$$\mu_w = \sqrt{\frac{p(1-p)}{n}} \simeq \sqrt{\frac{w(1-w)}{n-1}} \quad \text{pentru selecția întîmplătoare repetată} \qquad (7.22)$$

(7.23)

$$\mu_w = \sqrt{\frac{p(1-p)}{n}\cdot\left(1-\frac{n}{N}\right)} \simeq \sqrt{\frac{w(1-w)}{n-1}\cdot\left(1-\frac{n}{N}\right)}^{1} \quad \text{pentru selecția întîmplătoare nerepetată}$$

[1] Dacă volumul selecției este suficient de mare (peste 120 unități) se poate renunța la -1 din numitorul erorii medii cînd acest indicator se calculează pe baza dispersiei de selecție

În concluzie, se poate afirma că, în cazul respectării condiţiilor impuse de aplicarea unei selecţii aleatoare, calculul erorii medii de reprezentativitate se poate face *cu anticipaţie*, pe baza dispersiei caracteristicii studiate şi a volumului de selecţie, precizia estimaţiilor fiind în funcţie de probabilităţile de apariţie ale mediilor de selecţie.

7.3.1.3. EROAREA LIMITĂ DE REPREZENTATIVITATE

În practică, însă, la efectuarea unei singure selecţii poate să apară orice valoare a mediei cuprinse între nivelul său cel mai îndepărtat sau cel mai apropiat de media colectivităţii generale. La începutul capitolului s-a arătat că se consideră ca realizare a reprezentativităţii, selecţiile care nu produc erori mai mari de $\pm 5\%$. Aceasta înseamnă că este necesar să se limiteze intervalul de variaţie a mediei de selecţie, cu alte cuvinte, să se stabilească *mărimea erorii limită*. Eroarea limită de reprezentativitate se determină ca o abatere între media de selecţie şi media colectivităţii generale garantată de suma probabilităţilor corespunzătoare limitelor intervalului său de variaţie. După cum s-a arătat, la organizarea unei observări selective se porneşte de la ipoteza că media de selecţie este o variabilă aleatoare, care poate lua valori diferite şi că fiecărei valori îi corespunde o anumită frecvenţă, care se transformă în probabilitate de apariţie a mediei de selecţie şi că, în cazul efectuării tuturor eşantioanelor posibile, suma probabilităţilor de apariţie este egală cu 1.

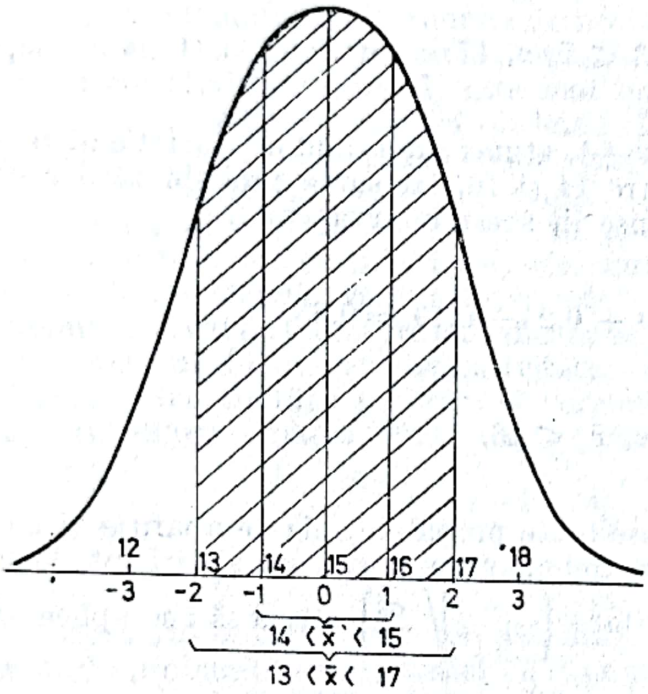


Fig. 7.2. Distribuţia erorilor de selecţie.

Grafic, ele se prezintă ca în figura 7.2.

Dacă se notează cu Δx — eroarea limită în cazul caracteristicii nealternative şi cu Δw — eroarea limită în cazul caracteristicii alternative se poate vedea că ele pot lua valori în ambele sensuri efectuîndu-se diferenţa între media de selecţie obţinută şi media colectivităţii generale. În mod corespunzător, media generală va fi egală cu media de selecţie plus sau minus eroarea limită ($\bar{x}_0 = \bar{x} \pm \Delta x$), garantată de suma probabilităţilor cuprinse între cele două limite ale intervalului de variaţie a erorii de reprezentativitate aleasă.

În exemplul considerat, pentru o eroare limită de ± 2, suma probabilităţilor va fi dată de probabilităţile ce aparţin intervalului în care va-

riază media de selecţie care produce eroarea respectivă, adică între 13 şi 17.

Acest interval este acoperit de suma probabilităţilor de apariţie a tuturor mediilor de selecţie cuprinse în acest interval, adică :

$$\sum_{i=1}^{5} P_i = 0{,}125 + 0{,}19 + 0{,}25 + 0{,}19 + 0{,}125 = 0{,}860.$$

Se poate afirma că în exemplul luat, din 100 de selecţii efectuate, în 86 de cazuri media de selecţie va aparţine acestui interval de valori, dînd o eroare de selecţie care să nu depăşească ± 2, iar media generală cu probabilitatea respectivă va fi :

$$13 \leqslant \bar{x}_0 \leqslant 17.$$

Dacă se limitează eroarea la ± 1, atunci intervalul de variaţie al mediei de selecţie va fi cuprins între 14 şi 16, iar suma probabilităţilor de apariţie a valorilor medii cuprinse în acest interval va fi :

$$\sum_{i=1}^{3} P_i = 0{,}19 + 0{,}25 + 0{,}19 = 0{,}63,$$

iar :

$$14 \leqslant \bar{x}_0 \leqslant 16.$$

Cum în practică nu se cunosc toate probabilităţile de apariţie şi nici toate mediile de selecţie posibile, trebuie să se pornească şi aici tot de la mărimea erorii de reprezentativitate $\left(\mu_x = \sqrt{\frac{\sigma_0^2}{n}}\right)$ căreia să i se aplice un coeficient (t), adică :

$$\Delta_x = t\mu_x = \begin{cases} \Delta_x = t\sqrt{\dfrac{\sigma_0^2}{n}} & \text{— selecţia întîmplătoare repetată ;} \quad (7.24) \\ \Delta_x = t\sqrt{\dfrac{\sigma_0^2}{n} \cdot \left(1 - \dfrac{n}{N}\right)} & \text{— selecţia întîmplătoare nerepetată} \quad (7.25) \end{cases}$$

Acest coeficient (t) se determină în raport cu probabilitatea cu care se garantează rezultatele, se găseşte gata tabelat (anexa 2) şi este argumentul funcţiei (t) dată de integrala :

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{\frac{-t^2}{2}} dt \qquad (7.26)$$

Cu cît se dau lui t valori din ce în ce mai mari, cu atît $\Phi(t)$ se apropie mai rapid de 1. De exemplu:

t	$\Phi(t)$	t	$\Phi(t)$
0,1	0,6827	3,4	0,9993
2,0	0,9545	3,8	0,99986
3,0	0,9973	4,0	0,999936

Se poate afirma, deci, că eroarea limită este şi ea o mărime variabilă, care depinde de coeficientul de probabilitate şi de eroarea medie a mediilor de selecţie. Rezultă că, pe lîngă modificarea volumului de selecţie ce intră în calculul erorii medii de reprezentativitate, eroarea limită de selecţie se poate modifica şi în funcţie de probabilitatea cu care se garantează apariţia ei. Dacă probabilitatea aleasă este mai mare, siguranţa estimaţiei mediei colectivităţii generale este mai mare ca urmare a creşterii intervalului de încredere a mediei obţinute. În mod corespunzător, abaterea mediei de selecţie de la media colectivităţii generale este mai mare şi, deci, exactitatea estimaţiei este mai mică.

Pentru a se putea menţine la un anumit grad de exactitate al estimaţiei, în practica statistică se poate formula problema şi invers: cu ce probabilitate trebuie garantate rezultatele selecţiei pentru a obţine un anumit interval de încredere a mediei de selecţie, stabilit în funcţie de mărimea erorii medii de reprezentativitate, calculată pe baza relaţiei dintre dispersia colectivităţii generale şi volumul selecţiei. În acest caz se face raportul dintre eroarea limită (Δ_x) şi eroarea medie (μ_x):

$$t = \frac{\Delta_x}{\mu_x}. \tag{7.27}$$

Se caută în tabele cărei probabilităţi îi corespunde mărimea coeficientului de probabilitate (t), astfel încît eroarea limită să se încadreze intervalului de valori care a fost stabilit.

În mod corespunzător se va calcula şi eroarea limită (Δw) pentru caracteristica alternativă, adică:

$$\Delta_w = t\mu_w = \begin{cases} \Delta_w = t\sqrt{\dfrac{p(1-p)}{n}} & \text{— pentru selecţia repetată;} \quad (7.28) \\ \Delta_w = t\sqrt{\dfrac{p(1-p)}{n}\left(1-\dfrac{n}{N}\right)} & \text{— pentru selecţia întîmplătoare nerepetată.} \quad (7.29) \end{cases}$$

iar coeficientul de probabilitate (t) va fi:

$$t = \frac{\Delta_w}{\mu_w}. \tag{7.30}$$

7.4. FELURILE SELECȚIILOR UTILIZATE FRECVENT ÎN PRACTICA STATISTICĂ. CALCULUL ERORILOR DE REPREZENTATIVITATE PE FELURI DE SELECȚIE

Pentru ca extinderea rezultatelor selecției la întreaga colectivitate să se facă în cît mai bune condiții erorile de reprezentativitate trebuie să fie minime. Pentru realizarea acestui scop trebuie să se cunoască care este gradul de omogenitate al colectivității studiate selectiv. Dacă se constată că este o colectivitate omogenă, atunci se poate folosi selecția întîmplătoare simplă — adică selecția se face prin extragerea individuală a unităților din întreaga colectivitate, pînă la completarea volumului eșantionului.

În cazul cînd colectivitatea este neomogenă, este necesar să se stabilească care sînt tipurile calitative care există în mod obiectiv în colectivitatea respectivă și fiecare să fie reprezentat în mod corespunzător în colectivitatea de selecție. În acest caz, deci, apare necesitatea grupării unităților pe tipurile calitative existente și apoi din fiecare grupă să se facă extragerea la întîmplare a unităților pentru a forma eșantionul. Din această cauză ea se numește *selecție tipică*. În literatura de specialitate selecția tipică se mai numește și selecție stratificată, în special cînd ea se efectuează din grupele populației, care se mai numesc și straturi ale populației.

La efectuarea în practică a selecției trebuie să se mai țină seama și de forma de organizare a colectivității ce formează obiectul cercetării. În cazul cînd ea este formată din unități de observare complexe, constituite sub forma unor serii de unități simple, reprezentativitatea eșantionului se realizează în măsura în care acesta reproduce trăsăturile esențiale ale tuturor seriilor ce formează colectivitatea generală, iar selecția pentru formarea eșantionului nu se mai face individual, ci extrăgînd cîte o serie întreagă de unități. Acest tip de selecție se numește *selecție de serii*.

Ținînd seama de toate aceste considerente, înseamnă că eroarea de reprezentativitate depinde nu numai de procedeul după care se aleg unitățile, dar mai ales de felul sau tipul de selecție utilizat. Mai frecvent în practica statistică se întîlnesc următoarele feluri de selecție, în funcție de care se modifică modul de calcul al erorii medii de selecție :

— selecția întîmplătoare simplă ;
— selecția mecanică ;
— selecția tipică ;
— selecția de serii ;
— selecția în trepte ;
— selecția în mai multe faze.

7.4.1. Selecția întîmplătoare simplă

Acest fel de selecție reprezintă forma de bază a aplicării metodei selective în cercetarea fenomenelor de masă și pe schemele ei sînt determinate formulele de calcul ale erorilor de selecție (vezi pag. 182—193). După cum s-a arătat, acest fel de selecție se aplică colectivităților omogene. În cazul în

Scanned with OKEN Scanner

care colectivitatea supusă cercetării este neomogenă, selecţia întîmplătoare simplă dă cele mai multe erori de reprezentativitate, deoarece dispersia colectivităţii generale măsoară întreaga variaţie din colectivitatea cercetată. De asemenea, cînd nu se cunoaşte gradul de omogenitate a colectivităţii care se cercetează prin aplicarea metodei selective, se foloseşte experimental selecţia întîmplătoare simplă, formîndu-se mai multe eşantioane, stabilind dispersia de selecţie pentru fiecare eşantion şi făcînd apoi media lor ca mijloc de verificare a omogenităţii colectivităţii respective.

$$\mu_{\bar{x}} = \sqrt{\frac{\overline{\sigma}}{n-1}}; \quad \mu_w = \sqrt{\frac{\overline{w(1-w)}}{n-1}}, \tag{7.31}$$

în care :

$\overline{\sigma}$, $\overline{w(1-w)}$ reprezintă media dispersiilor din eşantioanele cercetate.

Eroarea medie de reprezentativitate se poate calcula şi pe baza dispersiei colectivităţii generale calculată dintr-o cercetare anterioară sau se foloseşte dispersia de selecţie cînd colectivitatea studiată este suficient de omogenă, adică :

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma_0^2}{n}} \simeq \sqrt{\frac{\sigma_i^2}{n}}.$$

Această eroare medie este înmulţită cu coeficientul de corecţie a selecţiei nerepetate, dacă s-a folosit acest procedeu $\left(\sqrt{1-\frac{n}{N}}\right)$.

În selecţia întîmplătoare nerepetată erorile de reprezentativitate sînt mai mici decît cele obţinute în cazul selecţiei întîmplătoare simple repetate.

7.4.2. Selecţia mecanică

Selecţia mecanică, ca şi selecţia întîmplătoare simplă, se poate folosi şi ca procedeu de selecţie pentru alte feluri de selecţie — tipică sau de serii — şi ca tip de selecţie pentru formarea directă a eşantionului din întreaga colectivitate.

În literatura de specialitate selecţia mecanică este asimilată fie selecţiei întîmplătoare repetate, fie celei nerepetate. Din această cauză selecţia mecanică nu are formule proprii de calcul al erorilor de reprezentativitate. De regulă, se folosesc formulele de calcul de la selecţia repetată, care, după cum s-a văzut, conduc la erori de reprezentativitate mai mari decît în cazul selecţiei nerepetate. Ceea ce este specific pentru selecţia mecanică este pasul de măsuarare $\left(K = \frac{N}{n}\right)$, care stabileşte ordinea în care se includ în eşantion unităţile colectivităţii generale şi volumul colectivităţii de selecţie.

Selecţia mecanică se foloseşte în mod frecvent în agricultură, la determinarea recoltei medii probabile. Datorită faptului că fiecare probă se ia de pe o suprafaţă de 1 mp, cu ajutorul unei rame pătratice, cu latura de

Scanned with OKEN Scanner

1 m, este necesar să se determine distanţa de la care se aplică rama metrică, după formula :

$$d = \sqrt{\frac{S \cdot 10\,000}{n}}.$$

în care :

d reprezintă distanţa în metri ;
S — suprafaţa totală în ha ;
n — numărul probelor necesare pentru formarea eşantionului.

Calculînd eroarea mediei de selecţie pe baza formulelor de la selecţia repetată şi extinzînd rezultatele se obţine recolta medie probabilă pentru întregul lan.

7.4.3. Selecţia tipică (stratificată)

Este selecţia care se aplică cel mai frecvent în cercetarea fenomenelor sociale şi economice de masă. Aceste fenomene sînt, de regulă, fenomene cu o amplitudine mare a variaţiei, aparţinînd unor tipuri calitativ diferite, existente în mod obiectiv în colectivitatea studiată. Din această cauză, aceste colectivităţi, pentru a putea fi studiate, trebuie despărţite în grupe omogene. Formarea eşantionului, în acest caz, se va face aplicînd procedeele de eşantionare la fiecare grupă în parte. Aceasta face ca fiecare grupă să fie reprezentată în colectivitatea de selecţie, ceea ce asigură eşantionului un grad mai mare de reprezentativitate. Deci, se poate afirma că aplicarea selecţiei tipice este una din căile de reducere a erorii de reprezentativitate, deoarece grupele sînt omogene şi în cadrul lor sînt caracterizate de indicatorii de variaţie cu valori mici. Eroarea de selecţie, care depinde, în acest caz, de gradul de variaţie din fiecare grupă, va fi mai mică decît în cazul cînd se face selecţia din aceeaşi colectivitate nedespărţită în grupe tipice. Cu cît grupele în care a fost împărţită colectivitatea generală sînt mai omogene, cu atît mediile de grupă au valori mai apropiate de valorile individuale din care s-au calculat şi, deci, abaterile într-un sens şi altul sînt mai puţin semnificative. În acest caz, variaţia mediilor de selecţie posibile nu va mai depinde de gradul de variaţie din întreaga colectivitate (caracterizat prin mărimea dispersiei generale (σ_0^2), ci de gradul de variaţie a fiecărei grupe prezente în colectivitatea generală, sintetizat în mărimea mediei dispersiilor parţiale ($\overline{\sigma}_0^2$). Din relaţiile cunoscute, media dispersiilor parţiale este o parte componentă a dispersiei generale, adică :

$$\sigma_0^2 = \overline{\sigma}_0^2 + \delta_0^2,$$

deci :

$$\overline{\sigma}_0^2 < \sigma_0^2$$

Cu cît caracteristicile după care s-a făcut gruparea au un rol mai mare în formarea cîmpului de variaţie a caracteristicii studiate, cu atît mediile de grupă vor diferi mai mult între ele şi se vor abate într-o măsură mai mare de la media colectivităţii generale, ceea ce va face ca greutatea specifică

Scanned with OKEN Scanner

a dispersiei dintre grupe (δ_0^2) să fie mai mare în raport cu greutatea specifică a mediei dispersiilor parţiale ($\overline{\sigma^2}$).

Aceasta înseamnă că, în cazul selecţiei tipice, eroarea de reprezentativitate se calculează în raport cu mărimea mediei dispersiilor parţiale (de grupă), adică:

$$\mu_x = \sqrt{\frac{\overline{\sigma_0^2}}{n}} \simeq \sqrt{\frac{\overline{\sigma^2}}{n}} \quad \text{pentru selecţia tipică repetată;} \tag{7.32}$$

$$\mu_x = \sqrt{\frac{\overline{\sigma_0^2}}{n} \cdot \left(1 - \frac{n}{N}\right)} \simeq \sqrt{\frac{\overline{\sigma^2}}{n} \cdot \left(1 - \frac{n}{N}\right)} \quad \text{pentru selecţia tipică nerepetată.} \tag{7.33}$$

Ca în orice fel de selecţie, în cazul în care nu se dispune, dintr-o cercetare anterioară, de date obţinute dintr-o observare totală, în locul mediei dispersiilor parţiale din colectivitatea generală ($\overline{\sigma_0^2}$) se poate folosi media dispersiilor parţiale din colectivitatea de selecţie ($\overline{\sigma^2}$).

Selecţia tipică poate fi: simplă, proporţională şi optimă.

Selecţia simplă stratificată este aceea în care din fiecare grupă se extrage un număr de unităţi pentru a forma eşantionul, fără a ţine seama de greutatea specifică a fiecărei grupe în totalul colectivităţii.

Selecţia tipică proporţională este aceea în care fiecare grupă este reprezentată în eşantion în funcţie de greutatea sa specifică. În cazul selecţiei tipice proporţionale, structura colectivităţii de selecţie este aproximativ aceeaşi cu structura colectivităţii generale din punctul de vedere al grupelor în care a fost împărţită.

Selecţia tipică optimă este considerată aceea în care repartizarea pe grupe a volumului de selecţie este făcută ţinînd seama de variaţia din fiecare grupă a colectivităţii generale şi de locul pe care-l ocupă în structura colectivităţii generale.

Condiţia de bază a aplicării selecţiei stratificate optime este, deci, de a se cunoaşte abaterea medie pătratică din fiecare grupă, în funcţie de mărimea căreia se stabileşte fracţia de selecţie. Aceasta presupune ca, pe lîngă structura colectivităţii generale, să se cunoască, dintr-o observare totală anterioară, gradul de variaţie pentru caracteristica după care se face eşantionarea. De regulă, în acest caz trebuie să se folosească o grupare combinată.

Cel mai frecvent în practica statistică se foloseşte selecţia tipică proporţională, în care proporţia de selecţie $\left(\frac{n}{N}\right)$ se aplică la fiecare grupă în parte, cu condiţia ca să se aleagă o caracteristică esenţială de grupare. În general, acest tip de selecţie dă rezultate bune, deoarece fiecare grupă se regăseşte în acelaşi raport şi în colectivitatea de selecţie ca şi pe întregul ansamblu.

Pentru determinarea erorii medii şi a erorii limită, care pot să apară într-o selecţie tipică, se va lua un exemplu concret.

Se presupune că s-a organizat o observare selectivă cu privire la determinarea procentului mediu de îndeplinire al normelor şi a proporţiilor medii a celor care nu îndeplinesc planul, într-o întreprindere de 3 000

Scanned with OKEN Scanner

de muncitori. Pentru aceasta s-au împărțit muncitorii în trei grupe tipice după gradul de calificare și vechime și s-au selectat întîmplător și nerepetat cîte 10% din fiecare grupă. Datele înregistrate au fost sintetizate pe cele trei grupe tipice (tabelul 7.6).

Tabelul 7.6

Calculul mediei și abaterii medii pătratice în selecția tipică

Grupe tipice	Numărul muncitorilor în:		Media de selecție (%)	Proporția celor care n-au îndeplinit planul	Dispersia de selecție a gradului de îndepl. a normelor	Dispersia de selecție a celor care nu-și îndeplinesc planul
	Colectivitatea generală	Colectivitatea de selecție				
	N_i	n_i	$\bar{x}_i$	w_i	σ_i^2	$w(1-w)$
Gr .I	700	70	105	0,20	12,5	0,16
Gr. II	1 500	150	112	0,15	10,2	0,1275
Gr. III	800	80	120	0,10	8,4	0,09
	3 000	300	112,5	0,15	10,26	0,125

Media mediilor de selecție:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{r} \bar{x}_i n_i}{\sum_{i=1}^{r} n_i}$$

$$\bar{\bar{x}} = \frac{(105 \cdot 70) + (112 \cdot 150) + (120 \cdot 80)}{300} = 112{,}5\%$$

Media dispersiilor parțiale:

$$\bar{\sigma}^2 = \frac{(12{,}5 \cdot 70) + (10{,}2 \cdot 150) + (8{,}4 \cdot 80)}{300} = 10{,}26.$$

Eroarea medie de selecție:

$$\mu_{\bar{x}} = \sqrt{\frac{\bar{\sigma}^2}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{10{,}26}{300}\left(1 - \frac{300}{3\,000}\right)} = \pm 0{,}18.$$

Eroarea limită pentru o probabilitate de 0,999936 pentru care $t = 4$, va fi:

$$\Delta_{\bar{x}} = t\sqrt{\frac{\bar{\sigma}^2}{n}\left(1 - \frac{n}{N}\right)} = 4(\pm 0{,}18) = \pm 0{,}72.$$

Scanned with OKEN Scanner

Pentru a cunoaște proporția pe întregul eșantion a celor care nu și-au îndeplinit norma de producție va trebui să se calculeze media acestei caracteristici alternative :

$$\bar{w} = \frac{\sum_{i=1}^{r} w_i n_i}{\sum_{i=1}^{r} n_i} = \frac{(0{,}20 \cdot 70) + (0{,}15 \cdot 150) + (0{,}10 \cdot 80)}{300} = 0{,}15 \text{ sau } 15\%.$$

Eroarea medie de selecție pentru caracteristica alternativă, în cazul sondajului proporțional, va fi :

$$\mu_w = \sqrt{\frac{\overline{w_i(1 - w_i)}}{n}} \quad \text{pentru selecția tipică repetată}$$

$$\mu_w = \sqrt{\frac{\overline{w_i(1 - w_i)}}{n}\left(1 - \frac{n}{N}\right)} \quad \text{— pentru selecția tipică nerepetată.}$$

Pentru a calcula (μ_w) este necesar să se determine, mai întîi, mărimea dispersiei medii a caracteristicii alternative :

$$\overline{w_i(1 - w_i)} = \frac{\sum_{i=1}^{r} [w_i(1 - w_i)] n_i}{\sum_{i=1}^{r} n_i} =$$

$$= \frac{(0{,}16 \cdot 70) + (0{,}1275 \cdot 150) + (0{,}90 \cdot 80)}{300} = 0{,}126,$$

de unde :

$$\mu_w = \sqrt{\frac{0{,}125}{300}\left(1 - \frac{300}{3\,000}\right)} = \pm 0{,}01976 \simeq \pm 0{,}020.$$

Iar eroarea limită (Δ_w) pentru același coeficient de probabilitate :

$$\Delta_w = t\mu_w = 4(\pm 0{,}020) = \pm 0{,}080.$$

7.4.4. Selecția de serii

După cum s-a mai arătat, fenomenele sociale și economice, care se cercetează prin aplicarea metodei selective, mai diferă între ele și prin forma lor concretă de organizare. Existența obiectivă, în mod permanent, a unităților statistice sub forma unor unități complexe (echipe, brigăzi,

Scanned with OKEN Scanner

secţii, întreprinderi etc.) impune în mod necesar formarea eşantionului în funcţie de aceste forme de organizare a colectivităţii generale. Reprezentativitatea selecţiei constă, în acest caz, în asigurarea unei structuri asemănătoare a eşantionului, din punctul de vedere al formei de organizare a colectivităţii generale, sub formă de unităţi complexe. Pentru aceasta este necesar ca selecţia să nu se mai facă prin exragerea individuală a unităţilor simple, ci prin extragerea de unităţi complexe, care, în mod curent, se numesc serii. Din această cauză se numeşte selecţie de serii şi prezintă unele particularităţi în ceea ce priveşte asigurarea reprezentativităţii eşantionului şi a calculului erorii de selecţie.

În legătură cu primul aspect — asigurarea reprezentativităţii eşantionului — trebuie arătat că în cazul selecţiei de serii colectivitatea generală se prezintă împărţită, de regulă, în grupe eterogene. De exemplu, în cadrul unei brigăzi pot exista muncitori cu grade diferite de calificare, cu o vechime în producţie diferită, concretizată printr-un nivel de productivitate diferit. În acest caz, mediile din fiecare grupă depind de gradul variaţiei din cadrul grupei respective, asocierea factorilor esenţiali şi întîmplători producîndu-se fără nici o regularitate specifică fiecărei grupe. De regulă, în cadrul fiecărei grupe, modul de asociere a acestor factori urmează regularitatea specifică întregii colectivităţi. Dacă colectivitatea este formată din brigăzi, ele vor trebui să cuprindă muncitori din toate categoriile existente în întreprinderea respectivă, iar reprezentativitatea eşantionului va trebui să fie asigurată în funcţie de gradul de variaţie dintre seriile colectivităţii generale. Dacă, însă, seria se referă la un lot de piese produse, atunci ele vor forma unităţi relativ omogene, corespunzătoare omogenităţii întregului lot de produse. Ca urmare a acestui fapt mediile de grupă, calculate pe fiecare serie în parte, pot să se formeze ca valori foarte apropiate între ele şi, deci, apropiate de valoarea medie pe întreaga colectivitate. Rezultă, deci, că gradul de variaţie în colectivitatea generală depinde, în primul rînd, de dispersiile care există între seriile colectivităţii generale. Precizia rezultatelor selecţiei este în funcţie de măsura în care seriile colectivităţii generale sînt reprezentate în colectivitatea de selecţie. Aceasta înseamnă că aici nu se poate aplica selecţia individuală, ci se face extragerea de serii întregi pînă la completarea eşantionului.

Extrăgîndu-se pe rînd seriile pentru a forma colectivitatea de selecţie, înseamnă că fiecare serie trebuie să fie caracterizată printr-un anumit nivel de dezvoltare a caracteristicii. Aceasta nu se poate face decît dacă pentru fiecare serie extrasă se înregistrează toate nivelurile individuale şi se face media lor, media seriei apărînd ca o medie de grupă. Deci, distribuţia de selecţie se obţine aici ca o distribuţie de frecvenţe a mediilor de grupă, pe baza cărora trebuie calculată media lor. În acest fel, eroarea de selecţie apare ca o abatere dintre media seriilor care au intrat în selecţie şi media colectivităţii generale. Se poate spune că exactitatea estimaţiei mediei colectivităţii generale depinde de gradul de dispersie a mediilor calculate pe fiecare serie în parte şi, respectiv, de măsura în care mediile seriilor selectate se apropie de forma de distribuţie corespunzătoare seriilor din care este formată colectivitatea generală.

Din cele arătate reiese că eroarea medie de reprezentativitate depinde de gradul de variaţie dintre seriile colectivităţii generale, care se determină în mod sintetic prin dispersia dintre mediile calculate pe fiecare serie şi

media colectivității generale (σ_0^2). Cum și dispersia dintre serii este o parte componentă a dispersiei colectivității generale ($\delta_0^2 < \sigma_0^2$), înseamnă că și în acest caz erorile de reprezentativitate sînt mai mici sau cel mult egale cu cele care se produc cînd se efectuează o selecție întîmplătoare simplă, dintr-o colectivitate neîmpărțită în grupe. Deci, în acest caz în formula erorii limită va apare în locul dispersiei colectivității generale, dispersia dintre grupe notată cu δ_x^2 pentru caracteristica nealternativă și δ_p^2 pentru caracteristica alternativă.

Eșantionarea se face pe bază de serii, deci, selecția este formată dintr-un număr de serii notate cu r, iar pe întreaga colectivitate cu R.

Ținînd seama de toate aceste notații, eroarea medie de reprezentativitate va fi :

— pentru selecția repetată :

$$\mu_x = \sqrt{\frac{\delta_x^2}{r}} \text{ și respectiv } \mu_w = \sqrt{\frac{\delta_p^2}{r}}; \tag{7.34}$$

— pentru selecția nerepetată :

$$\mu_x = \sqrt{\frac{\delta^2}{r}\left(\frac{R-r}{R-1}\right)}; \quad \mu_w = \sqrt{\frac{\delta_p^2}{r}\left(\frac{R-r}{R-1}\right)} \tag{7.35}$$

Dacă se operează cu dispersia de selecție, eroarea medie este :

$$\mu_x = \sqrt{\frac{\delta^2}{r-1}\left(\frac{R-r}{R-1}\right)}; \quad \mu_w = \sqrt{\frac{\delta_w^2}{r-1}\left(\frac{R-r}{R-1}\right)}. \tag{7.36}$$

În mod corespunzător se determină și eroarea limită înmulțind eroarea medie (μ) cu coeficientul de probabilitate (t).

În cazul selecției de serii nerepetate, nu se mai renunță la -1 din numitorul coeficientului de corecție, deoarece lucrîndu-se cu serii, numărul lor este mult mai mic și neluarea lui în considerare afectează mărimea erorii de selecție calculată.

7.4.5. Alte feluri de selecție folosite în practica statistică

Varietatea fenomenelor sociale, particularitățile pe care le îmbracă în manifestarea lor concretă, fac ca, în practică, metoda selecției să se aplice diferențiat, în funcție de natura specifică a fenomenelor cercetate și a surselor de informație ce pot fi folosite. Problemele care se pun cu ocazia organizării cercetării selective devin cu atît mai numeroase, cu cît colectivitatea generală este mai numeroasă și mai variată în formele ei de manifestare. În aceste cazuri, pentru a ajunge la cunoașterea statistică a formelor individuale de manifestare este necesar ca selecția să se facă în mai multe trepte. Acest fel de selecție se numește *selecție în*

trepte sau *multistadială*. De exemplu, pentru a cerceta selectiv muncitorii din industrie, selecţia se poate realiza în mai multe trepte, ţinînd seama de forma de organizare a industriei, pe ramuri de producţie, pe întreprinderi, secţii, echipe.

În prima treaptă se pot selecta ramurile de producţie la care se va face observarea, în cea de a doua treaptă întreprinderile, iar în ultima treaptă echipele din cadrul acestora. Deci, în fiecare din cele trei trepte (stadii) s-au selectat grupe neomogene ce pot fi considerate ca serii de unităţi statistice. Această selecţie se prezintă, deci, ca o formă particulară a selecţiei de serii cu fracţie variabilă de selecţie, în care eroarea de selecţie depinde de dispersia dintre seriile fiecărui stadiu şi de proporţia de selecţie. În acest caz, seriile din prima treaptă poartă denumirea de serii primare, cele din treapta a doua serii secundare, în treapta a treia serii terţiale etc.

Formula *erorii medii de selecţie în mai multe trepte* va fi :

$$\mu = \sqrt{\frac{\delta_1^2}{r_1}(1 - f_1) + \frac{\delta_2^2}{r_2}(1 - f_2), \ldots}, \qquad (7.37)$$

în care :

$\delta^2_{1,2}\ldots$ reprezintă dispersiile dintre serii în fiecare treaptă ;

$f_1, f_2\ldots$ — proporţia de selecţie a diferitelor trepte (calculată ca raport între numărul seriilor selectate şi numărul total al seriilor din treapta respectivă) ;

$r_1, r_2\ldots$ — numărul seriilor extrase din fiecare fază.

În practică, se poate ca în una din stadiile iniţiale să se aleagă în selecţie toate seriile din stadiul respectiv (de exemplu, se iau întreprinderi din toate ramurile). În acest caz, eroarea de selecţie nu mai depinde de dispersia dintre seriile respective, deoarece ele au fost cuprinse în întregime în observare şi, deci, aici nu mai apar erori de reprezentativitate.

Tot o selecţie de trepte poate fi considerată şi *selecţia teritorială* în care judeţele reprezintă baza de selecţie a primei trepte. În treapta a doua se pot alege localităţile din judeţele respective, iar în treapta a treia întreprinderile. De menţionat este faptul că în ultima treaptă se poate face şi o selecţie individuală, de exemplu, în întreprinderi se aleg muncitorii după selecţia simplă sau selecţia tipică. Deci, selecţia în mai multe stadii se poate combina cu selecţia simplă, tipică sau de serii.

În practica statistică se pot folosi şi *selecţii în mai multe faze*. Astfel, se poate organiza o observare totală în care se înregistrează un număr redus de caracteristici statistice. În faza a doua se înregistrează un număr mai mic de unităţi, dar cu un program de observare mai bogat ; în faza a treia un număr mai mic de unităţi, dar cu o caracterizare mai completă a fenomenelor studiate. Desigur că se poate folosi observarea selectivă în toate fazele cercetării, pe baza aceluiaşi volum sau cu un volum diferit de la o selecţie la alta. Se poate folosi selecţia în mai multe faze şi atunci cînd se organizează pentru prima oară o cercetare statistică în domeniul respectiv şi pe baza datelor de selecţie, se studiază gradul de omogenitate a colectivităţii şi diferenţierea ei pe tipuri calitative.

Scanned with OKEN Scanner

În general, în selecţia în mai multe faze se foloseşte ca indicator de variaţie, în funcţie de care se determină eroarea medie pătratică, media dispersiilor din fiecare fază, adică :

$$\mu_x = \sqrt{\frac{\overline{\sigma^2}}{n-1}} \quad \mu_w = \sqrt{\frac{\overline{w(1-w)}}{n-1}}, \tag{7.38}$$

în care :

$\overline{\sigma^2}$, $\overline{w(1-w)}$ reprezintă mediile dispersiilor de selecţie obţinute în fiecare fază de observare.

În sfîrşit, în practică se mai pot folosi eşantioane fixe, cum sînt, de pildă, cele organizate în domeniul statisticii nivelului de trai, sub forma bugetelor de familie, sau în studiul pieţei, paneluri de cumpărători. În acest caz, selecţia s-a făcut iniţial pe baza unor principii de selecţie şi reprezentativitatea eşantionului se verifică în funcţie de mărimea dispersiilor corespunzătoare fiecărei perioade, păstrîndu-se acelaşi volum, sau schimbînd volumul şi structura eşantionului în funcţie de modificările intervenite în colectivitatea generală. În cazul selecţiei pe baza bugetelor de familie, selecţia capătă aspect de observare curentă a nivelului de trai al populaţiei. Pe baza datelor culese se poate urmări dinamica fenomenelor cercetate folosind în acest scop indicatorii de selecţie, sistematizaţi sub formă de serii cronologice. În mod corespunzător se pot întocmi astfel de serii de date şi din parametrii estimaţi pentru întreaga colectivitate.

Alegerea tipului de selecţie trebuie să se facă în funcţie de gradul de precizie a rezultatelor şi de mărimea cheltuielilor ce se fac cu observarea şi prelucrarea datelor de selecţie.

Ţinînd seama de gradul de precizie a rezultatelor selecţiei în estimarea parametrilor colectivităţii generale este necesar să se găsească cel mai ieftin tip de selecţie care să asigure gradul de siguranţă propus, sau se pleacă de la un anumit volum maxim al cheltuielilor şi atunci, în funcţie de aceasta, este necesar să se găsească acel tip de selecţie care satisface o precizie cît mai mare în estimarea parametrilor căutaţi.

Aceasta presupune, deci, găsirea maximului sau minimului funcţiei care exprimă legătura dintre volumul de cheltuieli specifice (c) necesare pentru folosirea unui anumit volum de selecţie (n) determinat pe baza unei erori de selecţie care să nu depăşească intervalul de încredere ales, adică :

$$C = c_1 n_1 + c_2 n_2 + \ldots = \min., \tag{7.39}$$

unde C reprezintă cheltuielile totale necesare de selecţie. Rezultă, deci, că unul din factorii determinanţi ai preciziei selecţiei este volumul colectivităţii de selecţie.

7.5. DETERMINAREA VOLUMULUI DE SELECŢIE

Stabilirea numărului unităţilor care vor intra în selecţie nu se face la întîmplare. Precizia rezultatelor selecţiei, posibilitatea extinderii lor asupra întregii colectivităţi depind şi de numărul de unităţi la care se face

Scanned with OKEN Scanner

culegerea datelor. Determinarea volumului de selecţie este condiţionată, în primul rînd, de precizia cu care trebuie estimaţi parametrii colectivităţii generale. După cum s-a arătat, gradul de precizie al estimaţiilor depinde de distribuţia probabilităţilor de apariţie a mediilor de selecţie, care şi ea depinde de mărimea şi forma de variaţie a caracteristicii în colectivitatea generală.

Deci, volumul colectivităţii de selecţie depinde de dispersia colectivităţii generale (σ_0^2), de probabilitatea cu care se garantează apariţia mediei de selecţie exprimată prin mărimea coeficientului de probabilitate (t) şi de intervalul de valori în care trebuie să plaseze eroarea de reprezentativitate (Δ_x şi, respectiv, Δ_w).

Toate aceste elemente, inclusiv volumul selecţiei care trebuie determinat, sînt cuprinse în formula de calcul al erorii limită.

Determinarea volumului de selecţie se face pe baza formulei de calcul al erorii limită pentru felul şi procedeul de selecţie care urmează să se folosească.

Astfel, în cazul selecţiei repetate pentru caracteristica nealternativă, va fi :

$$\Delta_x = t \cdot \sqrt{\frac{\sigma_0^2}{n}},$$

de unde, prin ridicare la pătrat, dispare radicalul, adică :

$$\Delta_x^2 = t^2 \frac{\sigma_0^2}{n} \quad \text{iar} \quad n = \frac{t^2 \sigma_0^2}{\Delta_x^2}, \tag{7.40}$$

iar pentru caracteristica nealternativă :

$$\Delta_w = t \sqrt{\frac{p(1-p)}{n}} \text{ în care } n = \frac{t^2 p(1-p)}{\Delta_w^2} \tag{7.41}$$

În cazul selecţiei nerepetate, la determinarea volumului de selecţie trebuie să ţinem seama şi de proporţia de selecţie $\left(\frac{n}{N}\right)$, respectiv :

— pentru caracteristica nealternativă :

$$\Delta_x = t \sqrt{\frac{\sigma_0^2}{n}\left(1 - \frac{n}{N}\right)}$$

prin ridicare la pătrat :

$$\Delta_x^2 = \frac{t^2 \sigma_0^2}{n} \cdot \frac{N-n}{N};$$

eliminînd numitorii :

$$n \cdot N\Delta_x^2 = Nt\sigma_0^2 - nt\sigma_0^2;$$

— izolînd necunoscuta n şi scoţînd-o ca factor se va obţine :

$$n(N\Delta_x^2 + t^2\sigma_0^2) = N \cdot t^2 \sigma_0^2,$$

Scanned with OKEN Scanner

de unde:

$$n = \frac{Nt^2\sigma_0^2}{N\Delta_x^2 + t^2\sigma_0^2}. \tag{7.42}$$

Pentru a putea fi reţinută mai uşor, prin analogie cu formula de calcul pentru selecţia repetată, se simplifică cu N fracţia din membrul doi, adică:

$$n = \frac{t^2\sigma_0^2}{\Delta_x^2 + \frac{t^2\sigma_0^2}{N}}. \tag{7.43.}$$

Folosind acelaşi interval de variaţie a erorii de reprezentativitate volumul eşantionului necesar pentru o selecţie nerepetată este mai mic decît cel calculat pentru selecţia repetată, deoarece numitorul este mai mare.

În mod corespunzător, pentru caracteristica alternativă volumul selecţiei întîmplătoare repetate va fi:

$$n = \frac{t^2 p(1-p)}{\Delta_w^2 + \frac{t^2 p(1-p)}{N}} \tag{7.44}$$

Pentru *selecţia tipică* volumul de selecţie se stabileşte în mod corespunzător, introducînd în formula de calcul în locul dispersiei colectivităţii generale media dispersiilor de grupă:

— pentru caracteristica nealternativă:

$$n = \frac{t^2\overline{\sigma_0^2}}{\Delta_x^2}; \quad n = \frac{t^2\overline{\sigma_0^2}}{\Delta_x^2 + \frac{t^2\overline{\sigma_0^2}}{N}}; \tag{7.45}$$

— pentru caracteristica alternativă:

$$n = \frac{t^2\overline{p(1-p)}}{\Delta_w^2}; \quad n = \frac{t^2\overline{p(1-p)}}{\Delta_w^2 + \frac{t^2\overline{p(1-p)}}{N}}. \tag{7.46}$$

În plus, la selecţia stratificată se pune problema repartizării volumului de selecţie pe grupe (n_1, $n_2 \ldots n_r$).

Dacă selecţia este stratificată proporţional, atunci greutatea specifică a fiecărei grupe trebuie să coincidă şi în eşantion şi în colectivitatea de bază. Deci, volumul eşantionului se repartizează pe subeşantioane în

Scanned with OKEN Scanner

funcţie de greutatea specifică a fiecărei grupe în colectivitatea generală, adică :

$$n_i = n \frac{N_i}{\sum_{i=1}^{r} N_i}. \tag{7.47}$$

Această relaţie se mai poate scrie :

$$n_i = n \frac{N_i}{\sum_{i=1}^{r} N_i} = \frac{n_i}{\sum_{i=1}^{r} N_i} \cdot N_i ; \text{ Deci } n_i = \frac{n}{N} \cdot N_i \tag{7.48}$$

Folosind această relaţie se stabileşte proporţia de selecţie $\left(\frac{n}{N}\right)$ şi aceasta se aplică numărului de unităţi din care este formată fiecare grupă a colectivităţii generale (N_i).

Dacă selecţia este stratificată optim, atunci trebuie să se stabilească volumul de selecţie pe grupe ţinînd seama atît de proporţia de selecţie, cît şi de gradul de variaţie din fiecare grupă măsurat prin mărimea abaterii sale medii pătratice (σ_i), după formula :

$$n_i = n \frac{N_i \sigma_i}{\sum_{i=1}^{r} N_i \sigma_i}. \tag{7.49}$$

Pentru exemplificare se vor folosi datele din tabelul 7.7., considerînd o colectivitate generală de 8 000 de unităţi (N) împărţită în trei grupe tipice $(r = 3)$ şi avînd abaterile medii pătratice (σ_i) indicate.

Tabelul 7.7

Calculul mediei dispersiilor parţiale

Grupele tipice	Volumul colectivităţii generale N_i	Abaterile medii pătratice de grupă în colectivitatea generală σ_i	Dispersiile de grupă în colectivitatea generală σ_i^2	$\sigma_i N_i$	$\sigma_i^2 N_i$
0	1	2	3	4	5
I	2 500	20,8	432,64	52 000	1 081 600
II	4 000	40,5	1 640,25	162 000	6 561 000
III	1 500	15,8	249,64	23 700	274 000
Total	8 000	—	—	237 700	7 917 060

Pentru a determina volumul eşantionului este necesar să se stabilească cu ce probabilitate se vor garanta rezultatele selecţiei şi între ce limite trebuie să se plaseze eroarea de reprezentativitate. Presupunînd că rezultatele selecţiei vor fi garantate cu o probabilitate de 0,999936 pentru

Scanned with OKEN Scanner

care ($t = 4$), iar eroarea maximă (Δ_x) într-un sens sau altul va fi de 5, volumul selecţiei se va calcula după formula :

$$n = \frac{t^2 \overline{\sigma_0^2}}{\Delta_x^2}.$$

În exemplul considerat media dispersiilor parţiale va fi :

$$\overline{\sigma_0^2} = \frac{\Sigma \sigma_i^2 N_i}{\Sigma N_i} = \frac{7\,917\,060}{8\,000} = 989{,}63.$$

Deci :

$$n = \frac{4^2 \cdot 989{,}63}{5^2} = 633{,}36 \simeq 640.$$

În continuare trebuie să se repartizeze acest volum pe grupe :

— pentru selecţia *tipică simplă* se repartizează cele 640 unităţi pe cele trei grupe, adică din fiecare grupă urmează să se selecteze cîte 213 unităţi $\left(\frac{640}{3} = 213\right)$;

— pentru selecţia *tipică proporţională :*

$$n_i = n \frac{N_i}{\sum_{i=1}^{r} N_i}$$

— gr. I $\quad n_1 = 640 \dfrac{2\,500}{8\,000} = 200$

— gr. II $\quad n_2 = 640 \dfrac{4\,000}{8\,000} = 320$

— gr. III $\quad n_3 = 640 \dfrac{1\,500}{8\,000} = 120$

$$n = 200 + 320 + 120 = 640.$$

Sau se stabileşte proporţia de selecţie $\left(\frac{n}{N}\right)$ care se aplică fiecărei grupe din colectivitatea generală. Dacă volumul colectivităţii de selecţie este de 640 şi volumul colectivităţii generale este de 8 000, atunci fracţia de selecţie este de $\frac{640}{8\,000} = 0{,}08$ sau 8%, iar pe grupe eşantionul se poate forma astfel :

gr. I $\quad n_1 = \dfrac{2\,500 \cdot 8}{100} = 200$;

gr. II $\quad n_2 = \dfrac{4\,000 \cdot 8}{100} = 320$;

gr. III $\quad n_3 = \dfrac{1\,500 \cdot 8}{100} = 120$;

Scanned with OKEN Scanner

— dacă *selecția este optim repartizată* atunci volumul fiecărei grupe se va obține astfel:

$$n_i = \frac{n\sigma_{i_0} N_i}{\sum_{i=1}^{r} \sigma_{i_0} N_i}$$

— gr. I $\quad n_1 = \frac{640 \cdot 52\,000}{237\,700} = 140$

— gr. II $\quad n_2 = \frac{640 \cdot 162\,000}{237\,700} = 436$

— gr. III $\quad n_3 = \frac{640 \cdot 23\,700}{237\,700} = 64$

$$n = 140 + 436 + 64 = 640.$$

Dacă se calculează acum fracția de selecție $\left(\frac{n_i}{N_i}\right)$ ea este variabilă de la o grupă la alta și față de întreaga colectivitate, care este de 8%.

gr. I $\quad \frac{140}{2\,500} \cdot 100 = 5{,}6\%;$

gr. II $\quad \frac{436}{1\,500} \cdot 100 = 10{,}9\%;$

gr. III $\quad \frac{64}{1\,500} \cdot 100 = 4{,}26\%.$

Deci, ținîndu-se seamă și de structura colectivității generale pe grupe și de gradul variației acestora, se poate afirma că grupa a doua, care este mai puțin omogenă — cu cea mai mare abatere medie pătratică — este mai bine reprezentată în colectivitatea de selecție, în timp ce grupa a treia de selecție, mai omogenă, este mai slab reprezentată (4,26%). Deoarece în cazul selecției optime proporția de selecție variază de la o grupă la alta, acest fel de selecție se mai numește și selecția cu fracția variabilă. La acest fel de selecție structura pe grupe a eșantionului nu corespunde cu structura colectivității generale ca în cazul selecției tipice proporționale.

Pentru comparație se vor calcula greutățile specifice ale fiecărei grupe în colectivitatea de selecție și în eșantionul proporțional și cel optim (tab. 7.8).

Este evident că, în acest caz, dacă se va folosi eșantionul optim, repartizat pe grupe, eroarea de selecție va fi mai mică decît în eșantionarea tipică proporțională, deoarece s-a asigurat o reprezentativitate mai mare a grupei care este cea mai numeroasă și cea mai puțin omogenă. Rezultă că selecția tipică optimă este indicat să fie folosită în cazul în care abaterile medii pătratice diferă substanțial de la o grupă la alta; altfel nu va diferi prea mult față de selecția tipică proporțională.

Scanned with OKEN Scanner

Tabelul 7.8

Repartizarea eşantionului pe grupe folosind selecţia proporţională şi optimă

Grupele	Colectivitatea generală		Eşantion proporţional		Eşantion optim	
	Nr. de unităţi	în %faţă de total	Nr. de unităţi	în% faţă de total	Nr. de unităţi	în % faţă de total
0	1	2	3	4	5	6
I	2 500	31,25	200	31,25	140	21,8
II	4 000	50,00	320	50,00	436	68,2
III	1 500	18,75	120	18,75	64	10,0
	8 000	100,00	640	100,00	640	100,0

Pentru selecţia de serii, volumul eşantionului se exprimă în număr de serii ce trebuie selectate (r), adică:

— pentru selecţia repetată de serii:

$$\Delta_x = t\sqrt{\frac{\delta_x^2}{r}}; \quad r = \frac{t^2\delta_x^2}{\Delta_x^2}; \tag{7.50}$$

— pentru selecţia nerepetată de serii:

$$\Delta_x = t\sqrt{\frac{\delta_x^2}{r}\cdot\frac{R-r}{R-1}};$$

— prin ridicare la pătrat dispare radicalul:

$$\Delta_x^2 = t^2\frac{\delta_x^2}{r}\cdot\left(\frac{R-r}{R-1}\right);$$

— eliminînd numitorul şi efectuînd calculele în membrul doi:

$$r(R-1)\Delta_x^2 = Rt^2\delta_x^2 - rt^2\delta_x^2;$$

— izolînd necunoscuta r:

$$r[(R-1)\Delta_x^2 + t^2\delta_x^2] = Rt^2\delta_x^2,$$

de unde:

$$r = \frac{Rt^2\delta_x^2}{(R-1)\Delta_x^2 + t^2\delta_x^2} \tag{7.51}$$

Pentru caracteristica alternativă, în mod analog se va determina r folosind în locul dispersiilor dintre serii pentru caracteristica nealternativă

Scanned with OKEN Scanner

(δ_x^2) același tip de dispersie calculată pentru caracteristica respectivă (δ_p^2), adică :

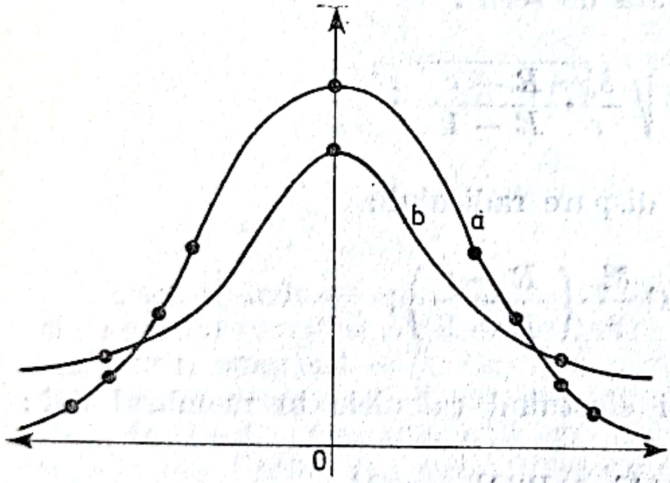
— la selecția repetată : la selecția nerepetată :

$$r = \frac{t^2\delta_p^2}{\Delta_w^2}; \qquad r = \frac{Rt^2\delta_p^2}{(R-1)\Delta_w^2 + t^2\delta_p^2}. \tag{7.52}$$

7.6. SELECȚIA DE VOLUM REDUS

În paragrafele precedente au fost tratate problemele care apar în legătură cu organizarea unei observări selective, în cazul unei selecții de volum normal, în care distribuția mediilor de selecție urmează la limită distribuția normală, indiferent de forma de distribuție din colectivitatea generală. Cunoscînd legea de distribuție a mediilor de selecție, eroarea de selecție se poate măsura în unități ale abaterii tip, cu probabilități de apariție a mediilor de selecție bine determinate, corespunzătoare funcției Laplace-Gauss.

Cînd volumul colectivității de selecție este mai mic (sub 40 de unități) forma de distribuție a mediilor de selecție depinde de forma de distribuție a colectivității generale. Cum, de regulă, forma de distribuție din colectivitatea generală nu este cunoscută, înseamnă că legea de distribuție a mediilor de selecție nu poate fi determinată. În teoria selecției se demonstrează că, în cazul selecției de volum mic, probabilitățile de apariție se distribuie mai uniform, adică ele nu descresc pe măsură ce se apropie de valorile extreme ale mediei de selecție. Deci, erorile de reprezentativitate mari au probabilități de apariție mai mari decît în cazul selecției de volum mare, mediile de selecție distribuindu-se după curba lui Student și nu după curba normală (vezi fig. 7.3).



a= distribuție normală
b= distribuție student

Fig. 7.3. Distribuția erorilor de selecție.

Deci, în cazul selecției de volum redus probabilitățile de apariție nu pot fi calculate pe baza tabelării distribuției Laplace-Gauss, ci pe baza tabelării Student, în care argumentul funcției depinde de distribuția de probabilități și de volumul de selecție.

Dacă se notează cu $S(z)$ coeficientul de probabilitate, atunci z va fi egal ca în cazul selecției de volum normal, cu raportul dintre eroarea limită (Δ_x) și eroarea sa medie (μ_x), adică :

$$z = \frac{\Delta_x}{\mu_x}, \tag{7.53}$$

în care z reprezintă argumentul funcției $S(z)$. Teoretic, între $\Phi(t)$ și $S(z)$ există o legătură și anume :

$$\Phi(t) = 2S(z) - 1, \tag{7.54}$$

de unde :

$$S(z) = \frac{\Phi(t) + 1}{2}. \tag{7.55}$$

Pe baza acestei relații și a tabelului de valori ale funcției z, se poate calcula și aici, cu anticipație, la ce eroare de selecție se poate aștepta, în cazul în care se folosește un volum redus al selecției.

Cum nu se dispune de dispersia colectivității generale, eroarea medie de selecție se calculează folosind dispersia de selecție, raportată la volumul micșorat cu un grad de libertate, adică :

$$\mu_x = \sqrt{\frac{\sigma^2}{n-1}} \text{ și } \mu_w = \sqrt{\frac{w(1-w)}{n-1}}. \tag{7.56}$$

Deoarece erorile de reprezentativitate care se pot obține sînt mai mari în cazul selecției de volum normal, chiar dacă selecția s-a făcut nerepetat, eroarea medie nu se corectează cu coeficientul de corecție $\sqrt{1 - \frac{n}{N}}$ și se folosesc formulele de calcul de la selecția repetată.

Eroarea limită se calculează diferențiat și anume :

— sub 40 de unități după formulele :

$$\Delta_x = z\sqrt{\frac{\sigma^2}{n-1}} \text{ și } \Delta_w = z\sqrt{\frac{w(1-w)}{n-1}}; \tag{7.57}$$

— între 40 și 120 de unități după formulele :

$$\Delta_x = t\sqrt{\frac{\sigma^2}{n-1}} \text{ și } \Delta_w = t\sqrt{\frac{w(1-w)}{n-1}}. \tag{7.58}$$

Pentru selecțiile sub 20 de unități, eroarea limită depinde atît de volumul selecției, cît și de probabilitatea cu care se garantează rezultatele, între 20—40 unități se lucrează cu valorile lui z, corespunzătoare coloanei ultime a tabelului (anexa 3).

Selecțiile de volum redus se folosesc, în special, la controlul calității produselor în timpul fabricației, pentru prevenirea eventualelor dereglări ale utilajelor.

Se presupune că într-o secție a unei întreprinderi metalurgice se finisează o piesă cilindrică pentru al cărui cap proiectul tehnic prevede un diametru de 5 cm, cu o toleranță în $\pm 0{,}15$ cm. Prin sondaj s-au măsurat 5 piese, stabilindu-se un diametru mediu de 5,06 cm, cu o dispersie de selecție de 0,00092[1]).

[1]) M. Biji, I. Stoichiță, *Metoda selectivă în cercetarea statistică* Ed. Științifică, București, 1957, 164 — 165.

Pornind de la aceste date, eroarea medie de selecţie va fi :

$$\mu_x = \sqrt{\frac{\sigma^2}{n-1}} = \sqrt{\frac{0{,}00092}{5-1}} = 0{,}01516.$$

Eroarea limită pentru o probabilitate de 0,9949 trebuie stabilită în funcţie de valoarea lui z, din tabelul distribuţiei Student pentru un volum de selecţie de 5 unităţi.

Pentru a afla probabilitatea corespunzătoare distribuţiei Student, se va determina pe $S(z)$ în raport cu $\Phi(t)$:

$$S(z) = \frac{\Phi(t) + 1}{2} = \frac{0{,}9949 + 1}{2} = 0{,}99745.$$

Căutînd în tabel pe coloana lui $n = 5$ nu se găseşte direct probabilitatea respectivă, se face interpolarea între $S(z) = 0{,}997$, căreia îi corespunde $z = 5{,}4$ şi $S(z) = 0{,}998$ pentru care $z = 5{,}6$. Prin interpolare $z = 5{,}49$.

De aceea eroarea limită va fi :

$$\Delta_x = z\sqrt{\frac{\sigma^2}{n-1}} = 5{,}49\sqrt{\frac{0{,}00092}{5-1}} = 0{,}085 \text{ cm}.$$

Deci, în cazul selecţiei de 5 unităţi, în 9 949 cazuri din 10 000, media de selecţie va avea o abatere de $\pm 0{,}085$ cm de la diametrul mediu pe întreaga colectivitate.

7.7. EXTINDEREA REZULTATELOR SELECŢIEI ASUPRA COLECTIVITĂŢII GENERALE

În practica statistică metoda selectivă se foloseşte, în principal, în două cazuri :

— ca mijloc de verificare a unei observări totale anterioare ;

— ca mijloc de a caracteriza statistic o colectivitate generală la care nu se poate aplica observarea totală.

Extinderea rezultatelor se poate realiza prin procedee diferite, în funcţie de cazurile amintite.

În primul caz se utilizează mai frecvent procedeul coeficientului de corecţie sau determinarea valorilor ecuaţiilor de regresie lineară între datele observării selective şi cele ale observării totale.

Procedeul coeficientului de corecţie constă în stabilirea unui raport între datele corespunzătoare pentru aceleaşi unităţi obţinute prin observarea totală. Acest coeficient se extinde apoi pentru toate unităţile la care s-a făcut observarea totală, înmulţind datele totalizate cu coeficientul respectiv. Acest procedeu se bazează pe faptul că, în cazul selecţiei, erorile de înregistrare sînt mai puţin numeroase şi mai uşor de evitat.

Scanned with OKEN Scanner

Se presupune că s-a organizat un recensămînt prin care s-au cules toate datele de la toate unităţile colectivităţii şi care au fost centralizate pe întreaga colectivitate. Ulterior s-a stabilit un eşantion de 15% şi s-au înregistrat aceleaşi caracteristici. Comparînd datele centralizate din selecţie cu cele centralizate pe baza înregistrărilor anterioare de la cele 15% unităţi care au intrat în selecţie, s-au obţinut următoarele valori :

Tabelul 7.9

Estimarea erorilor prin aplicarea coeficientului de corecţie

Denumirea caracteristicii	Date centralizate pe întreaga colectivitate	Date centralizate la 15% din unităţile de observare		Coeficient de corecţie	Date pe întreaga colectivitate corectate	Abateri între datele corectate şi cele obţinute prin observarea totală
		din observarea totală	din observarea selectivă			
0	1	2	3	4	5	6
x	500 000	80 000	84 000	1,05	525 000	+25 000
y	350 000	20 000	18 800	0.94	329 000	−21 000
⋮						

Procedeul pe baza extinderii ecuaţiilor de regresie presupune calcularea valorilor teoretice ale datelor observării totale (y) în funcţie de datele selecţiei (x) pentru aceleaşi unităţi, adică :

$$Y = a + bx. \tag{7.59}$$

Acest procedeu dă, în general, o mai mare siguranţă extinderii rezultatelor selecţiei asupra datelor obţinute din observarea totală anterioară.

Procedeul extinderii directe este cel mai frecvent utilizat în aplicarea metodei selective ca mijloc de caracterizare a întregii colectivităţi.

Pentru extinderea rezultatelor selecţiei se are în vedere, în primul rînd, determinarea intervalului de încredere în care să varieze media de selecţie, în funcţie de mărimea erorii sale limită, iar în al doilea rînd, cunoscînd volumul colectivităţii generale, determinarea limitelor de încredere ale nivelului totalizat al caracteristicii pe întreaga colectivitate.

Intervalul de încredere al mediei de selecţie este determinat de mărimea mediei de selecţie şi a erorii sale de reprezentativitate, adică :

$$\bar{x} - \Delta_{\bar{x}} < \bar{x}_0 < \bar{x} + \Delta_{\bar{x}}, \tag{7.60}$$

care, pe baza relaţiei lui Cebîşev, este dată de suma probabilităţilor ce acoperă mărimea intervalului respectiv, adică :

$P(|\bar{x} - \bar{x}_0| \leqslant t\mu) = \Phi(t)$ sau înlocuind cu formula sa :

$$P\left(|\bar{x} - \bar{x}_0| < t\sqrt{\frac{\sigma_0^2}{n}}\right) = \Phi(t). \tag{7.61}$$

Deci, pentru a cunoaşte intervalul de valori între care se va plasa media colectivităţii generale trebuie să se cunoască mărimea de selecţie şi mări-

Scanned with OKEN Scanner

mea erorii sale limită. Astfel, presupunînd că s-a organizat o cercetare selectivă privind mărimea medie a timpului care se pierde în cadrul unui schimb într-o întreprindere cu 3 000 muncitori, selecţia a cuprins 300 muncitori. În urma prelucrării datelor de selecţie timpul mediu care se pierde într-un schimb a fost de 22′, cu o eroare limită de ±2, calculată pentru $P = 0{,}9534$, pentru care $t = 3{,}20$.

Limitele intervalului de încredere al mediei de selecţie folosite pentru estimarea colectivităţii generale vor fi:

$$22' - 2' < \overline{x_0} < 22' + 2';$$

$$20' < \overline{x_0} < 24'.$$

Pe baza acestui interval de încredere se poate determina intervalul de valori în care se va plasa timpul total ce se pierde pe întreaga întreprindere în cadrul unui schimb de lucru, astfel:

$$(\bar{x} - \Delta_x) N < \sum_{i=1}^{N} x_i < (\bar{x} + \Delta_x) N \qquad (7.62)$$

$$20' \cdot 3\,000 < \sum_{i=1}^{3000} x_i < 24' \cdot 3\,000.$$

Deci, timpul care se pierde în cadrul întreprinderii într-un schimb este cuprins între 60 000′ — 72 000′. Transformînd în ore se poate afirma, cu un grad de siguranţă de 95,34 %, că cei 3 000 muncitori pierd într-un schimb între 1 000—1 200 ore.

În cazul caracteristicii alternative se pune problema să se determine frecvenţa relativă a acestei caracteristici în colectivitatea generală (p) pe baza proporţiei ei în selecţie (w) şi a erorii sale limită (Δ_w).

Pentru caracteristica alternativă, relaţia lui Cebîşev devine:

$$P\left(|w - p| < t\sqrt{\frac{p(1-p)}{n}}\right) = \Phi(t) \qquad (7.63)$$

Pe baza acestei relaţii intervalul de încredere al mediei caracteristicii alternative pe întreaga colectivitate va fi:

$$w - \Delta_w < p < w + \Delta_w. \qquad (7.64)$$

Dacă s-a organizat o observare selectivă în care proporţia de rebut a fost de 3,0%, cu o eroare limită de ±0,015 calculată pe baza unei probabilităţi de 0,866 pentru $t = 1{,}5$, atunci procentul mediu de rebut va fi:

$$0{,}03 - 0{,}015 < p < 0{,}03 + 0{,}015;$$

$$0{,}015 < p < 0{,}045.$$

Scanned with OKEN Scanner

Deci, procentul mediu de rebut se va încadra între 1,5%—4,5%, iar numărul absolut al rebuturilor în colectivitatea generală se va obține înmulțind intervalul de încredere al mediei de selecție cu volumul colectivității generale. Rezultă că, cu o probabilitate de 0,8663, numărul total al rebuturilor pe întregul lot va fi:

$$(w - \Delta_w)N < M < (w + \Delta_w)\,N; \tag{7.65.}$$

$$\frac{1{,}5 \cdot 10\,000}{100} < M < \frac{4{,}5 \cdot 10\,000}{100};$$

$$150 < M < 450.$$

În mod frecvent, în practica statistică, pentru studierea variației rebuturilor, interesează numai limita maximă a numărului și a frecvenței relative a acestora în întregul lot de produse. Există cazuri cînd acest coeficient maxim este estimat anticipat și atunci se pune problema ca prin sondaj să se verifice dacă această limită maximă a fost respectată. Cu alte cuvinte, este necesar să se verifice care este numărul absolut al rebuturilor care poate să fie admis, astfel încît să nu depășească procentul maxim calculat pe întregul lot de produse. Aceasta înseamnă să se calculeze care este *eroarea absolută a frecvenței relative în eșantion*, astfel încît să se încadreze limitelor admise ale procentului acceptat.

Pornind de la exemplul precedent se poate formula următoarea problemă: care este numărul maxim de piese rebut ce poate să-l conțină un eșantion format din 500 piese selectate întîmplător și repetat în așa fel, încît greutatea specifică a acestora pe întregul lot să nu depășească 4,5%, iar rezultatele să fie garantate cu o probabilitate de 0,8664 pentru care $t = 1{,}5$.

Pentru a rezolva această problemă trebuie să se pornească de la relația care există între proporția de rebut din eșantion (w) și cea din colectivitatea generală (p).

Dacă $w = p$, atunci eroarea frecvenței relative este egală cu zero, pe baza relației:

$$nw = np. \tag{7.66}$$

Dacă $w \neq p$, atunci mărimea diferenței dintre ele este egală cu Δ_w. Notînd eroarea frecvenței relative cu Δ_m, se poate calcula astfel:

$$\Delta_m = n\Delta_w; \text{ adică: } \Delta_m = nt\sqrt{\frac{p(1-p)}{n}}. \tag{7.67}$$

În exemplul luat eroarea limită a frecvenței calculate a fost de 0,015, iar volumul selecției ce urmează să se efectueze este de 500, deci:

$$\Delta_m = 500(\pm 0{,}015) = \pm 7{,}5 \text{ piese rebut.}$$

Scanned with OKEN Scanner

Pentru a obține numărul absolut al rebuturilor pe care poate să-l conțină eșantionul (m), astfel încît să nu depășească procentul maxim admis, se va deduce valoarea sa din formula erorii absolute a frecvenței relative în eșantion (Δ_m), adică :

$$\Delta_m = np - m; \text{ de unde : } m = np - \Delta_m. \tag{7.68}$$

Pentru un procent maxim de 4,5 %,numărul minim de rebuturi va fi :

$$m = \frac{500 \cdot 4{,}5}{100} - (500 \pm 0{,}015) = 15 \text{ piese.}$$

Deci, cu o probabilitate de 0,8664 se poate afirma că, în sondajul efectuat, cel mult 23 piese din 500 pot să fie rebut, astfel încît să se încadreze procentului admis pe întreaga colectivitate, producînd astfel o eroare maximă absolută de 8 piese. Dacă în sondajele succesive, de cîte 500 piese, aceste piese depășesc limitele admise atunci este necesar să se repete sondajul pentru a verifica dacă se poate admite sau respinge întregul lot de produse.

În acest capitol au fost expuse numai principalele probleme ce pot apărea cu prilejul organizării unei cercetări selective concrete, în care, pe baza datelor unei singure selecții, să se poată stabili limitele între care se vor plasa erorile de selecție cu care se estimează parametrii colectivității generale.

Practica statistică a dovedit că selecția poate da rezultate eficiente dacă este bine organizată și se respectă cu stricteţe, în toate etapele cercetării, principiile teoriei selecției.

Cercetarea selectivă este o cercetare complexă în care posibilitățile de analiză statistică au cîmp larg de aplicare, deoarece, bazîndu-se pe un număr mai mic de observații, poate folosi un număr mai mare de caracteristici pentru care se culeg date. Ea se poate aplica la determinarea mediilor, în analiza dispersională, în studiul corelației dintre fenomene, în analiza factorială bazată pe metoda indicilor. Toate aceste aspecte fac ca metoda selecției să fie larg folosită în statistica social-economică.

Succesele obținute în ultimele decenii în domeniul organizării diferitelor observări selective nu constau numai în creșterea frecvenței aplicării acestora, ci mai ales în îmbunătățirea conținutului metodologic și a tehnicii de efectuare a lor.

Datorită atenției ce i s-a acordat și continuă să i se acorde, metoda selectivă este cea mai avantajoasă și mai bine fundamentată dintre toate observările statistice parțiale.

Practica statistică a arătat că cercetarea selectivă, științific organizată, dă rezultate suficient de precise pentru a răspunde operativ multiplelor sarcini de cunoaștere, control și conducere planificată a vieții economice. Cercetarea selectivă ocupă un loc de frunte în aplicarea largă a metodelor și procedeelor moderne de calcul în studiul fenomenelor social-economice.

Scanned with OKEN Scanner

ELEMENTE DE ANALIZĂ DISPERSIONALĂ

Studierea fenomenelor şi proceselor sociale şi economice prin metode statistice presupune folosirea unor indicatori cu conţinut de valori sintetice, care să prezinte proprietatea de a fi reprezentative pentru cel mai mare număr de valori individuale din care s-au calculat.

Statistica dispune de procedee care permit să se verifice în ce măsură aceste valori calculate, de regulă, sub formă de mărimi medii sau ecuaţii de tendinţă, au caracter de valori tipice pentru grupa sau colectivitatea de fenomene la care se referă.

Printre aceste procedee se situează şi analiza dispersională, cu ajutorul căreia se studiază mărimea şi frecvenţa cu care valorile reale ale unei caracteristici statistice se abat de la valorile teoretice calculate, precum şi măsura în care aceste variaţii sînt dependente sau independente de factorul de grupare. Se poate spune, deci, că analiza dispersională se foloseşte de fiecare dată cînd valorile concrete ale unei caracteristici statistice înregistrate pentru studiul variaţiei se estimează prin mai multe valori teoretice calculate în funcţie de variaţia unei alte caracteristici sau a unor forme de organizare a colectivităţilor statistice. De exemplu, se poate studia variaţia retribuţiilor primite de muncitori în funcţie de gradul lor de calificare, sau se poate studia variaţia calităţii producţiei de la un lot de produse la altul. Este evident că rezultatele analizei dispersionale vor fi interpretate în mod diferit pentru cele două exemple prezentate. În primul caz se verifică în ce măsură gradul de variaţie a calificării muncitorilor determină variaţia retribuţiilor primite, iar în cel de-al doilea caz se verifică independenţa variaţiei calităţii produselor de lotul din care fac parte.

În sfîrşit, trebuie precizat că analiza dispersională se aplică, în special, cînd datele provin dintr-o cercetare selectivă, iar din cercetările anterioare se dispune de informaţii cu privire la forma de distribuţie a caracteristicilor în colectivitatea generală. Aceste caracteristici pot fi considerate fie ca variabile independente sau factoriale, fie ca variabile dependente sau rezultative folosind o grupare simplă sau combinată. În consecinţă, rezultatele analizei dispersionale obţinute prin prelucrarea datelor dintr-un eşantion se interpretează folosind principiile teoriei probabilităţilor cu privire la distribuţia erorilor de reprezentativitate.

Pentru a efectua, deci, analiza dispersională trebuie să se înregistreze variaţia unei caracteristici statistice condiţionată de unul sau mai mulţi factori de grupare. În acest sens, caracteristica a cărei variaţie se studiază se consideră ca variabilă rezultativă şi se notează cu y, iar caracteristicile după care se face gruparea datelor se consideră variabile independente sau factoriale şi se notează cu $x_1, x_2 \ldots x_n$.

În cazul unui singur factor de grupare se pune problema să se stabilească cît din variaţia totală este produsă de factorul de grupare şi cît se datorează celorlalţi factori neînregistraţi — cunoscuţi şi sub denumirea

Scanned with OKEN Scanner

de factori reziduali — consideraţi cu acţiune constantă în toate cazurile individuale studiate.

Variaţia totală se descompune, astfel în două părţi:

— variaţia produsă de influenţa factorului x;

— variaţia produsă de factorii neluaţi în considerare, priviţi ca factori reziduali (z).

Dacă se folosesc mai mulţi factori de grupare, variaţia explicată se descompune pe grade de influenţă corespunzătoare fiecărui factor, iar variaţia neexplicată este considerată ca variaţie reziduală rezultată din influenţa factorilor neînregistraţi. Dacă factorii înregistraţi nu acţionează independent unul de celălalt, atunci variaţia totală cuprinde şi o parte care este rezultatul interinfluenţei dintre factori.

De exemplu, folosind doi factori independenţi între ei, variaţia totală se descompune în trei părţi:

— variaţia produsă de factorul x_1;

— variaţia produsă de factorul x_2;

— variaţia produsă de factorii reziduali (z).

Dacă se înregistrează doi factori interdependenţi, atunci variaţia totală se poate descompune în patru părţi componente:

— variaţia produsă de influenţa factorului x_1;

— variaţia produsă de influenţa factorului x_2;

— variaţia produsă de interacţiunea dintre x_1 şi x_2;

— variaţia produsă de factorii reziduali (z).

În principal, analiza dispersională se aplică în următoarele cazuri:

— la verificarea dependenţei variaţiei unei caracteristici statistice de variaţia factorilor săi determinanţi;

— la verificarea independenţei variaţiei unei caracteristici statistice de criteriul de înregistrare sau de sistematizare a datelor individuale de masă;

— la verificarea stabilităţii mediei şi dispersiei de la o selecţie la alta.

În cazul cînd analiza dispersională se foloseşte pentru a verifica dependenţa variaţiei unei caracteristici statistice de variaţia factorilor săi determinanţi, ea este considerată ca o metodă auxiliară a metodei regresiei şi corelaţiei statistice (vezi capitolul „Analiza legăturilor dintre fenomenele social-economice"), folosindu-se atît înainte, cît şi după aplicarea regresiei.

Folosită înainte de aplicarea metodei corelaţiei, analiza dispersională serveşte, de fapt, pentru verificarea gradului de semnificaţie a factorului considerat principal pentru producerea variaţiei caracteristicii rezultative. Dacă din analiza dispersională, rezultă dependenţa variabilei rezultative de factorul înregistrat, atunci se trece la măsurarea gradului de interdependenţă prin metoda corelaţiei şi regresiei statistice. Dacă se foloseşte însă după aplicarea metodei regresiei şi corelaţiei statistice, ea are ca obiectiv verificarea veridicităţii funcţiei matematice cu ajutorul căreia s-au estimat valorile caracteristicii rezultative în funcţie de variaţia caracteristicii factoriale.

Dacă analiza dispersională este folosită pentru verificarea independenţei variaţiei caracteristicii de influenţa factorului de grupare, înseamnă că unităţile au fost grupate după un criteriu organizatoric (de exemplu; gruparea muncitorilor pe echipe; gruparea producţiei pe loturi etc.), care nu trebuie să influenţeze asupra variaţiei caracteristicii de la o grupă la

Scanned with OKEN Scanner

alta. În general, aceste grupe trebuie să corespundă cu structura calitativă a colectivității de bază. Dacă colectivitatea de bază este eterogenă — cum este cazul colectivității de muncitori — atunci echipele trebuie să aibă caracter eterogen, urmînd la limită distribuția din colectivitatea de bază. Dacă colectivitatea este omogenă — cum este cazul producției — atunci fiecare lot de produse trebuie să fie omogen, încadrîndu-se în limitele de toleranță admise. Cu alte cuvinte, în acest caz trebuie să se verifice, prin analiza dispersională, că variația caracteristicii care se studiază depinde numai de condițiile din interiorul grupelor și care sînt aceleași cu cele din colectivitatea generală, ceea ce înseamnă că nu există variații semnificative de la o grupă la alta.

Analiza dispersională se poate folosi și pentru verificarea egalității mediilor de la o selecție la alta, ceea ce ar corespunde, de fapt, cu păstrarea aceluiași grad de reprezentativitate a mai multor eșantioane succesive. Această problemă este tratată într-un cadru mai larg al verificării ipotezelor statistice cu privire la apariția și distribuția erorilor de la o selecție la alta.

În funcție de scopul aplicării analizei dispersionale se interpretează și rezultatele. Astfel, în primul caz dispersia provocată de factorul de grupare — factor determinant — trebuie să aibă valoarea cea mai mare, pentru a se putea confirma dependența statistică. În celelalte două cazuri dispersia provocată de factorul de grupare — factor organizatoric — trebuie să fie minimă, pentru a confirma gradul de interdependență a variației caracteristicii studiate de condițiile organizatorice.

8.1. INDICATORII UTILIZAȚI ÎN ANALIZA DISPERSIONALĂ

După cum s-a mai arătat, fenomenele supuse analizei statistice sînt fenomene complexe, care depind de un număr mare de factori, unii cu caracter esențial, manifestîndu-se în mod sistematic, alții cu caracter întîmplător, manifestînd o tendință de distribuție întîmplătoare. Cu cît factorul esențial acționează mai puternic, cu atît variația caracteristicii rezultative este mai dependentă de variația factorului ales. De aceea, aplicarea analizei dispersionale presupune interpretarea gradului și formei de variație a caracteristicii rezultative, în funcție de unul sau mai mulți factori esențiali. Aceasta înseamnă că ea se bazează pe metoda grupării, prin care unitățile observate se separă în grupe după variația caracteristicii de grupare, care este considerată drept factor de influență, și apoi se analizează care este dispersia termenilor în cadrul fiecărei grupe și între grupele ce au fost formate.

Cu cît factorul ales (x) este mai hotărîtor pentru formarea variabilei y, cu atît dispersia din interiorul fiecărei grupe este mai mică și mediile condiționate ($\bar{y}/x$) ale variabilei y în funcție de x sînt mai dispersate între ele și au, deci, abateri mai mari față de media colectivității totale ($\bar{y}_0$).

Variația totală a unei variabile aleatoare (y) este determinată de toți factorii de influență. Prin calcularea valorii medii se consideră că toți factorii esențiali și neesențiali au o acțiune constantă asupra variabilei (y) și, deci, s-ar înregistra aceeași valoare la toate unitățile, egală cu media colectivității totale ($\bar{y}_0$). Abaterile între valorile individuale (y_j) formate sub acțiunea tuturor factorilor și valoarea mediei colectivității totale ($\bar{y}_0$)

Scanned with OKEN Scanner

apar ca rezultat al acţiunii factorilor întîmplători, care pot acţiona în sens opus, şi potrivit legii numerelor mari, într-un număr suficient de mare de cazuri individuale a căror variaţie este întîmplătoare, acţiunea lor se compensează reciproc. Pentru a împiedica compensarea abaterilor, variaţia totală a caracteristicii (y) se măsoară pe baza pătratelor tuturor abaterilor individuale ale variantelor de la media lor. Dacă s-a înregistrat un singur factor de grupare, atunci variaţia totală se descompune în două părţi: variaţia produsă de factorul de grupare şi variaţia reziduală (variaţia factorilor neînregistraţi).

Pentru interpretarea variaţiei unei caracteristici statistice — după cum s-a arătat în capitolul „Mărimile medii, indicatorii variaţiei şi asimetriei" — se pot calcula indicatori absoluţi şi indicatori relativi. Pentru efectuarea analizei dispersionale, cel mai frecvent se folosesc următorii indicatori:

— devianţa (varianţa) termenilor de la media lor (Δ^2);
— gradul de determinaţie a factorilor de grupare (η^2);
— coeficientul raportului dintre dispersii (F).

Devianţa (varianţa) totală trebuie să arate, în mărime absolută, care este cîmpul de împrăştiere a tuturor valorilor individuale înregistrate. Considerînd ca valoare tipică pentru întreaga distribuţie media generală, înseamnă că abaterile tuturor valorilor individuale faţă de această valoare medie pot fi considerate ca măsuri ale influenţei tuturor factorilor esenţiali şi neesenţiali, care determină caracteristica a cărei variaţie se studiază.

Dacă se notează cu y_j variabila la care se aplică analiza dispersională, cu valorile individuale de la $y_1 \ldots y_m$, atunci varianţa totală se va calcula astfel:

$$\Delta_y^2 = \sum_{j=1}^{m} (y_j - \bar{y}_0)^2, \tag{8.1}$$

în care:

Δ_y^2 reprezintă devianţa totală a variabilei y[1)].

---

1) Acelaşi rezultat se obţine dacă abaterile se înlocuiesc cu formula binomului la pătrat, şi apoi media se înlocuieşte cu formula sa de calcul. De exemplu, pentru devianţa totală se obţine:

$$\Delta_y^2 = \sum_{j=1}^{m} (y_j - \bar{y}_0)^2 = \sum_{j=1}^{m} (y_j^2 - 2y_j\bar{y} + \bar{y}^2)$$

$$\Delta_y^2 = \sum_{j=1}^{m} \left(y_j^2 - 2\bar{y} \sum_{j=1}^{m} y_j + \sum_{j=1}^{m} \bar{y}^2 \right)$$

$$\Delta_y^2 = \sum_{j=1}^{m} y_j^2 - 2 \frac{\sum_{j=1}^{m} y_j}{n} \sum_{j=1}^{m} y_j + n \left(\frac{\sum_{j=1}^{m} y_j}{n} \right)^2.$$

Deci:

$$\Delta_y^2 = \sum_{j=1}^{m} y_j^2 - \frac{\left(\sum_{j=1}^{m} y_j \right)^2}{n}.$$

În mod analog se obţin şi celelalte formule pentru $\Delta_{y/x}^2$ şi $\Delta_{y/z}^2$.

Dar devianța totală, în cazul dependenței variabilei y de un factor x, se poate descompune în devianța pe seama factorului înregistrat $\Delta^2_{y/x}$ și devianța pe seama factorilor neînregistrați $\Delta^2_{y/z}$. Pentru a calcula influența factorului x, se calculează mediile condiționate notate cu $\bar{y}/x$ sau mai simplu $\bar{y}_i$. Dacă sînt r grupe, devianța pe seama factorului înregistrat se calculează prin însumarea abaterilor mediilor condiționate ($\bar{y}_i$) de la media colectivității totale, ridicate la pătrat, ponderate cu frecvența totală a fiecărei grupe:

$$\Delta^2_{y/x} = \sum_{i=1}^{r} (\bar{y}_i - \bar{y}_0)^2 n_i, \tag{8.2}$$

în care:

n_i reprezintă frecvența condiționată a fiecărei grupe formată în funcție de mărimea variabilei x.

Dacă grupele sînt egale, n_i poate fi scos în factor și se obține:

$$\Delta^2_{y/x} = n_i \sum_{i=1}^{r} (\bar{y}_i - \bar{y}_0)^2. \tag{8.3}$$

Devianța provocată de factorii neînregistrați se notează cu $\Delta^2_{y/z}$ și se calculează prin însumarea patratelor abaterilor termenilor individuali de la media lor de grupă:

$$\Delta^2_{y/z} = \sum_{i=1}^{r} \sum_{j=1}^{m} (y_j - \bar{y}_i)^2. \tag{8.4}$$

Devianța caracteristicii, fiind calculată pe baza abaterilor individuale, depinde de unitățile concrete de măsură în care este exprimată caracteristica respectivă. Pentru a compara rezultatele între ele este necesar să se calculeze indicatorii relativi. Pentru aceasta se calculează coeficienții gradului de determinație a factorilor înregistrați. Complementar se calculează coeficientul de nedeterminație.

Pentru a stabili gradul de determinare a factorului de grupare, care se notează cu η^2_x, se face raportul dintre devianța provocată de factorul considerat și devianța totală a caracteristicii rezultative:

$$\eta^2_x = \frac{\Delta^2_{y/x}}{\Delta^2_y}, \tag{8.5}$$

iar gradul de nedeterminare se notează cu η^2_z și se calculează ca raport între devianța reziduală și devianța totală a caracteristicii rezultative:

$$\eta^2_z = \frac{\Delta^2_{y/z}}{\Delta^2_y}. \tag{8.6}$$

Fiind valori relative complementare, suma lor este egală cu unitatea:

$$\eta^2_x + \eta^2_z = 1. \tag{8.7}$$

Scanned with OKEN Scanner

Cu cît valoarea coeficientului de determinaţie este mai mare, apropiindu-se de unitate, cu atît factorul de grupare — factor înregistrat — are o acţiune mai hotărîtoare asupra variabilei.

Pentru a măsura gradul de veridicitate a factorului de grupare cel mai frecvent se foloseşte coeficientul lui Fisher—coeficientul F — care se obţine ca raport între cele două dispersii corectate cu numărul gradelor de libertate. Pentru a o deosebi de dispersia calculată faţă de totalul unităţilor observate, dispersia corectată se va nota cu s^2.

$$F_{calc} = \frac{s^2_{y/x}}{s^2_{y/z}} \tag{8.8}$$

Dispersia corectată, care măsoară variaţia factorului înregistrat ($s^2_{y/x}$), se calculează raportînd devianţa provocată de factorul de grupare ($\Delta^2_{y/x}$) la numărul grupelor micşorat cu o unitate :

$$s^2_{y/x} = \frac{\sum_{i=1}^{r} (\bar{y}_i - \bar{y}_0)^2 n_i}{r-1} = \frac{\Delta^2_{y/x}}{n_x}, \tag{8.9}$$

în care :

r reprezintă numărul grupelor după variaţia lui x.

Dispersia corectată, calculată pentru variaţia provocată de factorii neînregistraţi, se calculează raportînd devianţa provocată de aceşti factori ($\Delta^2_{y/z}$) la numărul unităţilor înregistrate micşorat cu numărul grupelor :

$$s^2_{y/x} = \frac{\sum_{i=1}^{r} \sum_{j=1}^{m} (y_j - \bar{y}_i)^2}{N-r} = \frac{\Delta^2_{y/z}}{n_z}. \tag{8.10}$$

Rezultatul raportării celor două dispersii corectate se compară cu valoarea teoretică a acestui raport stabilit în funcţie de gradele de libertate luate în calcul. Distribuţia F se găseşte gata tabelată (vezi tabelul anexă 4) pentru diferite grade de libertate ale celor două dispersii calculate. Această valoare — criteriul F — reprezintă gradul de la care începe să fie semnificativă legătura dintre cele două variabile.

Formulele prezentate se referă la cazul în care datele sînt înregistrate în mod individual în fiecare grupă în parte. Cînd datele se grupează combinat după ambele caracteristici (x şi y) — cazul cînd există o amplitudine mare a variaţiei variabilei y — formulele se adaptează unui tabel cu dublă intrare. Se consideră factorul de grupare variabila x, şi se obţin distribuţiile condiţionate pentru variabila y (tabelul 8.1).

Devianţa totală se calculează pe baza distribuţiei marginale a caracteristicii y, după formula :

$$\Delta^2_y = \sum_{j=1}^{m} (y_j - \bar{y}_0)^2 f_j. \tag{8.11}$$

Scanned with OKEN Scanner

Tabelul 8.1

Distribuţiile variabilei y în funcţie de variaţia unei caracteristici factoriale x

Grupele colectivităţii după caracteristica factorială	Valorile caracteristicii y						Frecvenţe parţiale	Medii parţiale ale variabilei y $(\bar{y}/x_i = \bar{y}_i)$
	y_1	$y_2 \dots$		$y_j \dots$		y_m		
x_1	f_{11}	f_{12}	$\dots$	f_{1j}	$\dots$	f_{1m}	n_1	$\bar{y}_1$
x_2	f_{21}	f_{22}	$\dots$	f_{2j}	$\dots$	f_{2m}	n_2	$\bar{y}_2$
.	.	.		.		.	.	.
.	.	.		.		.	.	.
.	.	.		.		.	.	.
x_i	f_{i1}	f_{i2}	$\dots$	f_{ij}	$\dots$	f_{im}	n_i	$\bar{y}_i$
.	.	.		.		.	.	.
.	.	.		.		.	.	.
.	.	.		.		.	.	.
x_r	f_{r1}	f_{r2}	$\dots$	f_{rj}	$\dots$	f_{rm}	n_r	$\bar{y}_r$
	f_1	f_2	$\dots$	f_j	$\dots$	f_m	$\sum_{i=1}^{r} n_i = \sum_{j=1}^{m} f_j$	$\bar{y}_0$

Devianţa provocată de factorul x se calculează pe baza distribuţiei mediilor condiţionate ale caracteristicii y, după formula :

$$\Delta^2_{y/x} = \sum_{i=1}^{r} (\bar{y}_i - \bar{y}_0)^2 n_i. \tag{8.12}$$

Devianţa provocată de factorii neînregistraţi şi care acţionează numai în interiorul grupelor se calculează pe baza distribuţiilor condiţionate ale caracteristicii y însumîndu-se pătratele abaterilor pentru toate grupele formate după variaţia factorului de grupare, adică :

$$\Delta^2_{y/z} = \sum_{i=1}^{r} \sum_{j=1}^{m} (y_j - \bar{y}_i)^2 f_{ij}. \tag{8.13}$$

Între cele trei devianţe există aceeaşi relaţie, adică :

$$\Delta^2_y = \Delta^2_{y/x} + \Delta^2_{y/z} \tag{8.14}$$

După datele de care se dispune, sau pentru simplificarea calculelor, se pot folosi şi aici în locul abaterilor elementele de calcul de la momentele centrate de ordinul 2 adaptate la distribuţiile de frecvenţe.

8.2. APLICAŢII ALE ANALIZEI DISPERSIONALE ÎN CAZUL UNEI GRUPĂRI SIMPLE

Pentru a arăta tehnica de calcul a analizei dispersionale, se va folosi numai o grupare simplă, după variaţia factorului de grupare. Pentru o grupare combinată intervin în plus, înainte de însumare, efectuarea produselor de frecvenţă corespunzătoare abaterilor înregistrate : $(y_j - \bar{y}_i)$; $(\bar{y}_i - \bar{y}_0)$; $(y_j - \bar{y}_0)$.

Scanned with OKEN Scanner

Variația producției de tulpini în funcție de perioada de vegetație la cînepa de fibre

	x_i = perioada de vegetație în zile						Total
	55	70	85	100	115	130	
y_j producția de tulpini q/ha	12 10 14 15 13	18 20 24 22 19	25 27 30 29 26	39 36 34 40 37	48 46 44 54 52	64 57 66 59 62	
Suma perioadei	64	103	137	186	244	308	$\Sigma y_j = 1\,042$

Sursa: Prof. dr. doc. N. Ceapoiu, *Metode statistice aplicate în experiențele agricole și în biologie*, Ed. Agrosilvică București, 1968 (p. 175–176, tabelul 12).

Pentru a putea aprecia dacă perioada de vegetație este un factor determinant în producția de tulpini de cînepă, dat fiind cîmpul mare de variație $y_{min} = 10$ și $y_{max} = 62$, este necesar să se stabilească mediile condiționate pe fiecare grupă, formate după valoarea lui x.

De exemplu:

$$\bar{y}_{x=55} = \frac{12 + 10 + 14 + 15 + 13}{5} = \frac{64}{5};$$

$$\bar{y}_{x=55} = 12{,}8 \text{ ș.a.m.d.}$$

Mediile condiționate pe grupe sînt prezentate în continuare:

x	55	70	85	100	115	130	Total
$\bar{y}_x$	12,8	20,6	27,4	37,2	48,8	61,6	$\bar{y}_0 = 34{,}7$

Pentru a trece la efectuarea analizei dispersionale se vor calcula abaterile față de media condiționată, stabilită pe fiecare grupă și față de media colectivității totale (tab. 8.3).

Folosind datele din tabel, se obțin direct devianța totală (Δ^2_y) și devianța reziduală ($\Delta^2_{y/x}$) astfel:

$$\Delta^2_y = \sum_{j=1}^{m} (y_j - \bar{y}_0)^2 = 8\,501{,}90;$$

$$\Delta^2_{y/x} = \sum_{i=1}^{r} \sum_{j=1}^{m} (y_j - \bar{y}_0)^2 = 200{,}20.$$

Pentru a calcula devianța provocată de factorul de grupare:

$$\Delta^2_{y/x} = \sum_{i=1}^{r} (\bar{y}_i - \bar{y}_0)^2 n_i$$

Scanned with OKEN Scanner

Tabelul 8.3

Calculul variației factorilor neînregistrați și al variației totale a producției de tulpini de cînepă

Grupele după perioada de vegetație	Producția la ha (y_j)	$(y_j-\overline{y}_t)$	$(y_j-\overline{y}_t)^2$	$(y_j-\overline{y}_0)$	$(y_j-\overline{y}_0)^2$
Gr. I	12	−0,8	0,64	−22,6	515,29
	10	−2,8	7,84	−24,7	610,09
	14	+1,2	1,44	−20,7	428,49
	15	+2,2	4,84	−19,7	388,09
	13	+0,2	0,04	−21,4	470,89
Total gr. I	64	—	14,80	−109,5	2 412,85
Gr. II	18	−2,6	6,76	−16,7	278,89
	20	−0,6	0,36	−14,7	216,09
	24	3,4	11,56	−10,7	114,49
	22	1,4	1,96	−12,7	161,29
	19	−0,6	2,56	−15,7	246,49
Total gr. II	103	—	23,20	−70,5	1 017,25
Gr. III	25	−2,4	5,76	−9,5	94,09
	27	−0,4	0,16	−7,7	52,29
	30	2,6	9,76	−4,7	22,09
	29	1,6	2,56	−5,7	32,49
	26	−1,4	1,96	−8,7	75,69
Total. gr. III	137	—	17,20	−36,5	283,65
Gr. IV	39	1,8	8,24	4,3	18,49
	36	−1,2	1,44	1,3	1,69
	34	−3,2	10,24	−0,7	0,49
	40	2,8	7,84	5,3	28,09
	37	−0,2	0,04	2,3	5,29
Total gr. IV	176	—	22,80	—	54,05
Gr. V	48	−0,8	0,64	13,3	176,89
	46	−2,8	7,84	11,3	127,69
	44	−4,8	23,04	9,3	86,49
	54	5,2	27,04	19,3	372,49
	52	3,2	10,24	17,3	299,29
Total gr. V	244	—	68,80	—	1 062,85
Gr. VI	64	2,4	5,76	29,3	858,49
	57	−4,6	21,16	22,3	497,29
	66	4,4	19,36	31,3	979,69
	59	−2,6	6,76	24,3	590,49
	62	0,4	0,16	27,3	745,29
Total gr. VI	308	—	53,40	—	3 671,25
Total general	1 042	—	200,20	—	8 501,90
	$\sum_{j=1}^{m} y_j$	—	$\sum_{i=1}^{r}\sum_{j=1}^{m}(y_j-\overline{y}_t)^2$	—	$\sum_{j=1}^{m}(y_j-\overline{y}_0)^2$

se stabilesc abaterile dintre mediile condiționate și media colectivității totale, adică:

$$\Delta^2_{y/x} = (12{,}8 - 34{,}7)^2 5 + (20{,}6 - 34{,}7)^2 5 + (27{,}4 - 34{,}7)^2 5 +$$
$$+ (37{,}2 - 34{,}7)^2 5 + (48{,}8 - 34{,}7)^2 5 + (61{,}6 - 34{,}7)^2 5 = 8\,301{,}70.$$

Pe baza rezultatelor obținute coeficientul de determinație este:

$$\eta^2_x = \frac{8\,301{,}70}{8\,501{,}90} \cdot 100 = 97{,}6\%.$$

Deci se poate afirma că în eșantionul folosit 97,7% din variația totală este provocată de variația produsă de factorul de grupare (perioada de vegetație), restul de 2,3% fiind influența relativă a celorlalți factori. Pentru a stabili în ce măsură factorul de grupare poate fi considerat semnificativ, se face raportul dintre cele două dispersii. Cele două dispersii corectate sînt:

$$s^2_{y/x} = \frac{\sum_{i=1}^{r} (\bar{y}_i - \bar{y}_0)^2 n_i}{n_x} = \frac{8\,301{,}70}{5} = 1\,660{,}3;$$

$$s^2_{y/z} = \frac{\sum_{i=1}^{r} \sum_{j=1}^{m} (y_j - \bar{y}_i)^2}{n_z} = \frac{200{,}20}{24} = 8{,}34.$$

Cum era de așteptat, dispersia cea mai mare ca valoare este aceea provocată de factorul de grupare. Aceasta corespunde, de fapt, ipotezei că pentru întreaga producție de tulpini la cînepa de fibre se manifestă o tendință de distribuție normală a producției în funcție de perioada de vegetație cuprinsă între 55—130 zile.

Încadrîndu-se în cerințele legii de distribuție normală, se pot verifica datele experimentale cu ajutorul *criteriului lui Fisher*, care este elaborat în funcție de distribuția probabilităților pentru un nivel de semnificație de la 1% și 5%, luînd în considerare gradele de libertate cu care s-au obținut dispersiile corectate.

Pentru exemplul luat:

$$F = \frac{s^2_{y/x}}{s^2_{y/z}} = \frac{1\,660{,}3}{8{,}34} = 199.$$

Pentru a verifica gradul de semnificație a factorului de grupare se compară coeficientul F calculat cu cel tabelar. Dacă $F_{calc.} \geqslant F_{tab.}$ atunci legătura este semnificativă și caracteristica după care s-a efectuat gruparea este într-adevăr determinantă pentru variația caracteristicii rezultative. Pentru o probabilitate de 0,99, se caută în tabelul de distribuție F și pentru $n_1 = 5 (n_1 = n_x)$ și $n_2 = 24$ $(n_2 = n_z)$ se găsește o valoare a lui $F = 3{,}90$. Deci, $F_{0.99} = 3{,}90$ pentru $n_1 = 5$ și $n_2 = 24$ este mai mic decît coeficientul F calculat pe baza datelor observate (199). Deci se poate spune aproape cu certitudine că, pe baza datelor experimentale, producția de tulpini depinde, în principal, de perioada de vegetație la cînepa de fibre și pentru verificarea stabilității acestei legături se pot folosi datele de la mai multe selecții succesive pentru verificarea reprezen-

Scanned with OKEN Scanner

tativităţii eşantionului. Interpretarea probabilistică a fenomenelor presupune date reprezentative, în care distribuţia erorilor de estimaţie a diferiţilor parametri ai colectivităţii totale să se poată face cu anticipaţie pe baza unor valori calculate pentru distribuţiile teoretice şi tabelate. Din compararea rezultatelor obţinute, din datele observate, cu valoarea teoretică tabelară, se stabileşte gradul de semnificaţie a indicatorilor statistici calculaţi. Cu cît valoarea coeficientului F calculat este mai mare în comparaţie cu cel tabelar, cu atît factorul de grupare este mai semnificativ şi se poate trece la aplicarea metodelor de măsurare şi interpretare a corelaţiilor.

Analiza dispersională prezintă în plus avantajul că dă o anumită caracterizare numerică a gradului de intensitate a legăturii dintre caracteristica factorială şi cea rezultativă şi care va fi precizată prin calcularea indicatorilor sintetici de corelaţie obţinuţi prin metodele analitice de calcul statistic.

Pentru a înţelege mai clar succesiunea operaţiilor impuse de efectuarea analizei dispersionale se vor înscrie indicatorii folosiţi în tabelul 8.4.

Tabelul 8.4

Tabelul de calculul necesar efectuării analizei dispersionale în funcţie de un singur factor

Natura variaţiei	Devianţa (varianta) caracteristicii	Gradul de influenţă (%)	Numărul gradelor de libertate	Dispersiile corectate	F calculat	F tabelat
— între grupe (influenţa factorului de grupare x)	$\Delta^2_{y/x} = \sum_{i=1}^{r} (\bar{y}_i - \bar{y}_0)^2 n_i$	$\eta^2_x = \frac{\Delta^2_{y/x}}{\Delta^2_y}$	$n_x = r - 1$	$s^2_{y/x} = \frac{\Delta^2_{y/x}}{n_x}$	$F_{calc.} = \frac{s^2_{y/x}}{s^2_{y/z}}$	$F_{tabelar}$ $P = 0{,}95$ $P = 0{,}99$
— în interiorul grupelor (influenţa factorilor reziduali z)	$\Delta^2_{y/z} = \sum_{i=1}^{r} \sum_{j=1}^{m} (y_j - \bar{y}_i)^2 f_{ij}$	$\eta^2_z = \frac{\Delta^2_{y/z}}{\Delta^2_y}$	$n_z = (\Sigma f) - r$	$s^2_{y/z} = \frac{\Delta^2_{y/z}}{n_z}$		
Total	$\Delta^2_y = \sum_{j=1}^{m} (y_j - \bar{y}_0)^2 f_j$	$\eta^2_x + \eta^2_z = 1$	$(\Sigma f) - 1$	—	—	—

În încheiere trebuie subliniat că analiza dispersională nu poate fi considerată numai ca o tehnică de calcul statistic. Ea trebuie aplicată cu mult discernămînt şi numai la cazurile în care există garanţia că eşantionarea s-a făcut respectînd cu rigurozitate regulile impuse de teoria probabilităţilor cu privire la includerea aleatoare a unităţilor în eşantion

Practica demonstrează că se obţin rezultate semnificative numai în cazul în care distribuţia erorilor înregistrate prezintă tendinţa de normalitate. Pentru distribuţiile asimetrice compensarea abaterilor se face numai pe total, deoarece nu există o proporţionalitate a frecvenţelor. Rezultatele analizei dispersionale sînt folosite, în continuare, la interpretarea interdependenţei statistice.

Scanned with OKEN Scanner

CAPITOLUL IX

ANALIZA LEGĂTURILOR DINTRE FENOMENELE SOCIAL-ECONOMICE

Utilizarea din ce în ce mai frecventă a metodei corelaţiei în statistica socialistă este justificată de necesitatea crescîndă a reflectării într-o formă numerică adecvată a interdependenţei obiective dintre fenomenele social-economice în ceea ce priveşte natura, direcţia şi gradul de intensitate a legăturilor care se manifestă într-o anumită perioadă de timp sau în dinamică.

Materialismul istoric demonstrează în mod ştiinţific că fenomenele sociale apar şi se dezvoltă ca urmare a unor cauze variate, care pot acţiona în acelaşi sens sau în sens opus şi cu grad diferit de intensitate. Complexitatea interacţiunii dintre fenomene este cu atît mai mare, cu cît ele aparţin unei forme superioare de mişcare a materiei. Reiese, deci, că fenomenele sociale nu sînt, de regulă, fenomene univoc determinate, fiind rezultatul conjugării influenţei a multor fenomene-cauză, iar în sistemul acesta de legături nu toate raporturile de dependenţă au aceeaşi importanţă, acţiunea unora dintre ele compensîndu-se reciproc.

De aceea, în analiza statistică a raporturilor de dependenţă dintre fenomene, problema care se pune este aceea a măsurării relaţiei care există între două sau mai multe caracteristici cuprinse în programul unei cercetări concrete a fenomenelor social-economice de masă. Aceasta presupune, în primul rînd, să se constate dacă între caracteristica x — denumită caracteristica factorială sau independentă — şi caracteristica y — caracteristică rezultativă sau dependentă — există sau nu un raport de dependenţă şi, în al doilea rînd, dacă această relaţie există să se exprime printr-un indicator simplu sau sintetic de corelaţie măsura în care caracteristica factorială x contribuie la formarea caracteristicii rezultative y sub aspectul naturii, direcţiei şi formei de legătură dintre ele.

În acest stadiu de cercetare statistică analiza calitativă ce trebuie să preceadă aplicarea uneia sau a alteia dintre metodele statistice este deosebit de importantă. Aceasta permite ca, din ansamblul de legături existente în mod obiectiv între fenomenele aceleiaşi colectivităţi statistice, să se desprindă acelea care au caracter permanent şi sînt determinante în formarea nivelurilor concrete de dezvoltare a caracteristicii cercetate. Acest lucru apare cu atît mai necesar, cu cît legitatea de apariţie şi dezvoltare a fenomenelor social-economice fiind valabilă pentru întregul ansamblu, nu este sesizabilă, de obicei, în cadrul cercetării empirice, decît dacă datele se referă la un număr mare de cazuri individuale concrete, diferite ca forme de manifestare.

Avînd în vedere toate aceste aspecte, metoda corelaţiei nu poate da rezultate bune decît dacă se lucrează cu un număr suficient de mare de cazuri individuale şi în care distribuţia abaterilor este aproximativ normală. Dacă această condiţie nu este satisfăcută, cîmpul de acţiune a legii numerelor mari este limitată, iar concluziile desprinse în urma obţinerii anumitor

indicatori de corelaţie pot da naştere la o interpretare eronată, ca urmare a reflectării neveridice a fenomenelor supuse cercetării statistice. Şi cu această ocazie este necesar ca procesul de analiză să pornească de la simplu la complex, de la fenomene la esenţă, de la o esenţă de un anumit grad la o esenţă de grad superior. Aceasta presupune ca, în analiza legăturilor dintre fenomene, să se folosească metoda abstractizării succesive a factorilor, prin care să se poată studia atît legăturile simple, imediate dintre două fenomene legate printr-o relaţie de cauzalitate, cît şi legăturile parţiale şi multiple, prin care se evidenţiază atît relaţiile de cauzalitate directă, cît şi interacţiunea dintre factori.

Studiul statistic al interdependenţei dintre fenomene necesită identificarea legăturilor studiate de la cauză la efect, precum şi legăturile realizate prin intermediul unui şir de cauzalităţi. În acest sens este necesar ca relaţiile de cauzalitate din interiorul fenomenelor complexe să fie studiate şi prezentate tot sub formă de tendinţă valabilă la nivelul întregului ansamblu şi nu la nivelul unor valori individuale izolate. Aceasta conduce în mod obligatoriu la folosirea unor metode statistice în care să se ţină seama de formele de distribuţie de frecvenţă ale fenomenelor pentru care se studiază interdependenţa dintre ele şi care nu pot fi interpretate decît pe baza indicatorilor medii şi a celor de variaţie.

9.1. TIPURI DE LEGĂTURI ÎNTRE FENOMENELE SOCIAL-ECONOMICE

Formele de manifestare a relaţiilor de interdependenţă sînt extrem de variate şi adesea destul de greu de sesizat. Pentru a le studia este necesar să fie clasificate în funcţie de unele criterii, după care se pot deosebi unele de altele.

După natura relaţiei de cauzalitate legăturile dintre fenomene pot fi legături funcţionale şi legături statistice sau stohastice.

Legăturile funcţionale sînt univoce, realizate direct între un fenomen-cauză şi un fenomen-efect. Deci, fenomenul-efect depinde de o singură cauză, care poate fi identificată de cîte ori se produce. Rezultă că dacă condiţiile rămîn constante, atunci unei valori a caracteristicii factoriale îi corespunde o singură valoare a caracteristicii rezultative.

Relaţia matematică dintre fenomenul-efect şi fenomenul-cauză, pentru legăturile de tip funcţional este:

$$y = f(x).$$

Un exemplu de astfel de legătură funcţională este aceea dintre nivelul productivităţii muncii şi cheltuiala de timp de muncă. Economia politică demonstrează că pe măsură ce scade timpul de producere a unei mărfi, în aceeaşi măsură creşte şi productivitatea muncii.

Legăturile statistice (*stohastice*) sînt cele în care fenomenul-efect este rezultatul combinării influenţei mai multor cauze, care pot acţiona în condiţii egale sau în condiţii diferite. În această relaţie de multicauzalitate

Scanned with OKEN Scanner

unii factori au caracter esenţial, alţii întîmplător. Cu cît relaţiile de cauzalitate sînt mai numeroase, cu atît gradul de variabilitate al fenomenului-efect este mai mare.

În astfel de cazuri, la fiecare valoare a caracteristicii factoriale îi poate corespunde o distribuţie de valori a caracteristicii rezultative.

Relaţia matematică care exprimă o legătură de tip stohastic (statistic) este :

$$y = f(x_1, x_2, x_3, \ldots x_n), \qquad (9.1)$$

în care :

$x_1, x_2, x_3, \ldots x_n$ sînt caracteristicile-factori care au fost înregistrate şi determină variaţia caracteristicii rezultative y.

Specific legăturilor statistice este faptul că în variaţia unei caracteristici rezultative există întotdeauna şi o componentă aleatoare, care apare ca rezultat al interacţiunii dintre factorii esenţiali şi cei întîmplători.

Legăturile de tip stohastic sînt cele mai frecvente în domeniul fenomenelor sociale şi economice. În cazul legăturii statistice, identificînd un factor de influenţă, se poate constata că, pe măsura variaţiei acestuia, variază într-o măsură mai mare sau mai mică şi caracteristica care se consideră ca fiind dependentă de factorul ales. De exemplu, legătura dintre înzestrarea tehnică a muncii şi productivitatea muncii este o legătură de tip statistic. La o creştere a înzestrării tehnice a muncii se obţine o creştere diferită a productivităţii muncii, deoarece productivitatea muncii depinde şi de alţi factori. Pentru a studia gradul de dependenţă a productivităţii muncii de gradul de înzestrare tehnică este necesar ca aici să se folosească metoda abstractizării, în care să se considere că numai acest factor ar fi esenţial şi variabil, iar ceilalţi factori ar avea acţiune comună cu caracter constant, chiar dacă, în realitate, ei au o influenţă hotărîtoare asupra variaţiei productivităţii muncii.

Trebuie remarcat că legăturile de tip statistic pot fi şi reciproce, adică efectul se transformă, la rîndul lui, în cauză imediată sau mediată prin intermediul unor relaţii de cauzalitate în lanţ.

Pentru a studia legăturile de tip statistic este necesar *să se identifice* şi *să se ierarhizeze* factorii esenţiali, precum şi formele sub care se manifestă relaţiile de cauzalitate. Aceasta este posibil numai dacă se înregistrează toate unităţile care formează colectivitatea de fenomene ce depind de aceleaşi cauze esenţiale.

Cînd analiza relaţiilor de cauzalitate se studiază pe baza unor observări parţiale este necesar să se verifice, în prealabil, gradul de reprezentativitate al colectivităţii de selecţie şi să se verifice apoi gradul de semnificaţie al indicatorilor de corelaţie care s-au calculat. Şi aici interpretarea relaţiilor de cauzalitate, folosind datele de selecţie, se face în sens probabilist.

Varietatea formelor de manifestare a legăturilor statistice necesită în continuare o clasificare a lor după mai multe criterii. Un prim criteriu este acela al numărului factorilor înregistraţi.

După numărul caracteristicilor-factori luate în studiu, legăturile statistice pot fi : legături simple şi legături multiple.

Scanned with OKEN Scanner

Legăturile simple sînt acelea în care caracteristica rezultativă se studiază numai în funcţie de o singură caracteristică independentă, iar celelalte caracteristici-factoriale, chiar dacă au fost identificate şi înregistrate, se consideră cu acţiune constantă în toate cazurile individuale înregistrate.

Legăturile multiple presupun să se studieze dependenţa unei caracteristici rezultative în funcţie de mai mulţi factori înregistraţi. Interpretarea statistică a legăturilor multiple implică şi analiza legăturilor simple dintre toate caracteristicile înregistrate pentru calculul corelaţiei multiple.

După conţinutul caracteristicilor incluse în analiza de corelaţie, legăturile pot fi : de asociaţie şi de corelaţie.

Asociaţia statistică exprimă relaţia de interdependenţă dintre două sau mai multe caracteristici exprimate calitativ sau între o caracteristică calitativă şi una numerică. De exemplu, între aptitudini şi profesia aleasă există o legătură de tip stohastic, sau între gradul de îndemînare şi productivitatea muncii.

La folosirea caracteristicilor calitative este necesar să se găsească o posibilitate de cuantificare, pentru a putea trece apoi la calculul indicatorilor de corelaţie.

Corelaţia statistică exprimă relaţia de interdependenţă dintre două sau mai multe caracteristici exprimate numeric şi se poate măsura prin indicatori statistici de corelaţie. De exemplu, între nivelul de productivitate a muncii, vechimea în producţie şi nivelul retribuţiilor există legături de corelaţie, care pot fi analizate atît ca legături simple, cît şi ca o legătură multiplă.

După direcţia în care se produc, legăturile pot fi : directe şi inverse.

Legăturile directe sau în acelaşi sens se produc atunci cînd, pe măsură ce se modifică nivelul de dezvoltare al caracteristicii factoriale, se modifică în acelaşi sens şi nivelul caracteristicii rezultative. În exemplul precedent, atît productivitatea muncii, cît şi vechimea în producţie influenţează în acelaşi sens variaţia retribuţiilor.

Legăturile inverse sînt acelea în care, pe măsură ce se modifică nivelul de dezvoltare al caracteristicii factoriale, se modifică în sens contrar nivelul caracteristicii rezultative. De exemplu, între nivelul productivităţii muncii şi nivelul costului unitar există o legătură statistică inversă.

După forma legăturii, ele pot fi : *rectiliniare*, exprimate prin ecuaţia funcţiei liniei drepte şi *curbiliniare*, exprimate prin ecuaţia unei funcţii exponenţiale, parabolice, hiperbolice etc.

Identificarea formei de realizare a legăturii este determinantă pentru măsurarea corectă a corelaţiei dintre fenomenele social-economice.

După timpul în care se realizează, legăturile statistice pot fi : concomitente (sincrone) şi cu decalaj (asincrone).

Legăturile sincrone (concomitente) sînt cele care pot fi urmărite în dinamică pentru aceeaşi perioadă de timp. De exemplu, dacă se analizează corelaţia dintre dinamica productivităţii muncii şi a retribuţiilor, se poate observa că pe măsură ce creşte productivitatea muncii, creşte şi mărimea retribuţiilor încasate de muncitorii aceleiaşi colectivităţi statistice.

Legături asincrone (cu decalaj) apar atunci cînd caracteristicile factoriale încep să acţioneze asupra variaţiei caracteristice rezultative, după scurgerea unei perioade de timp. De exemplu, între mărimea investiţiilor în maşini şi instalaţii şi creşterea productivităţii muncii există un decalaj

Scanned with OKEN Scanner

necesar asimilării în masă a noilor tehnologii, sau între dezvoltarea unei ramuri noi de producţie şi mărimea exportului există un decalaj corespunzător asigurării competitivităţii produselor pe plan internaţional.

Problemele care trebuie rezolvate în aplicarea metodelor de analiză a corelaţiilor statistice pot fi sintetizate astfel :

— identificarea şi ierarhizarea factorilor care determină în mod obiectiv variaţia caracteristicii rezultative ;

— verificarea gradului de cuprindere a unităţilor înregistrate. Dacă unităţile observate provin dintr-o cercetare parţială de tip selectiv trebuie ca la interpretarea rezultatelor să se ţină seama de principiile teoriei probabilităţilor ;

— sistematizarea datelor observate, astfel încît să nu se modifice gradul şi forma de variaţie a caracteristicilor la care se aplică metoda corelaţiei ;

— verificarea existenţei şi formei de legătură dintre caracteristicile corelate, în vederea alegerii corecte a procedeelor statistico-matematice de măsurare a dependenţei statistice ;

— calcularea adecvată a indicatorilor de corelaţie în funcţie de forma de legătură şi de natura informaţiei de care se dispune ;

— aplicarea testelor de semnificaţie a indicatorilor de corelaţie pentru cazul în care ei provin dintr-un sondaj statistic.

9.2. METODE ŞI PROCEDEE DE VERIFICARE ŞI ANALIZĂ A LEGĂTURILOR STATISTICE

Aplicarea metodei corelaţiei pentru cercetarea interdependenţei dintre fenomenele şi procesele sociale şi economice, trebuie să ţină seama de forma obiectivă în care apar şi se dezvoltă legăturile care urmează să fie studiate, precum şi de posibilitatea reflectării lor prin expresii numerice adecvate.

Luînd în considerare primul aspect — cunoaşterea formei obiective de manifestare a legăturii — aceasta se realizează printr-o analiză logică materialist-dialectică, bazată pe cunoştinţele teoretice ale disciplinelor de specialitate din domeniul respectiv şi verificată prin metode de calcul şi interpretare statistică.

În ceea ce priveşte cel de-al doilea aspect — posibilitatea reflectării legăturilor statistice prin expresii numerice adecvate — aceasta depinde de natura specifică a fenomenelor care intră în raportul de interdependenţă şi de posibilitatea de măsurare a acestora, respectiv de natura informaţiei de care se dispune şi se poate folosi în măsurarea şi interpretarea acestor legături.

Ţinînd seama de aceste două aspecte, la cercetarea legăturilor dintre fenomene se pot folosi atît metode simple de interpretare a legăturii, cît şi metode analitice, bazate pe aplicarea procedeelor statisticii matematice.

În principal, metodele de analiză a corelaţiei dintre fenomenele de masă se pot clasifica în două grupe :

— metode şi procedee cu ajutorul cărora se pot constata existenţa sau lipsa corelaţiei ; verifica direcţia în care ea se realizează şi permit o

Scanned with OKEN Scanner

apreciere vizuală a tendinţei de manifestare a intensităţii legăturii dintre fenomenele supuse corelaţiei. Pentru acestea, cel mai frecvent se pot folosi : metoda seriilor statistice interdependente ; metoda grupărilor ; metoda grafică ; tabelul de corelaţie ; tabelul de asociere ; metoda balanţelor ; analiza dispersională ;

— metode şi procedee analitice de calcul şi interpretare statistică prin care se măsoară existenţa şi direcţia legăturii, precum şi forma şi gradul de intensitate în care se realizează. Aceste metode şi procedee permit, deci, exprimarea raporturilor de interdependenţă dintre fenomene, printr-un sistem de indicatori de corelaţie, printre care cei mai folosiţi sînt ecuaţiile de regresie, coeficienţii de corelaţie simplă, parţială şi multiplă, coeficienţii de corelaţie a rangurilor, coeficienţii de asociere.

Trebuie subliniat, încă o dată, că toate aceste metode şi procedee de calcul şi interpretare a corelaţiei nu se pot aplica decît dacă, prin cunoştinţele de care se dispune, s-a stabilit că în mod logic între fenomenele cercetate pot exista raporturi obiective de interdependenţă. Aceasta este cu atît mai necesar cu cît fenomenele social-economice pot prezenta, în general, o anumită simetrie în covariaţia lor în jurul valorilor medii fie sub aspectul gradului de împrăştiere a abaterilor, fie sub aspectul direcţiei producerii legăturii. Desigur că această relaţie de covariaţie nu este suficientă pentru a afirma că între caracteristicile respective există şi legături de corelaţie. Pentru ca o anumită covariaţie numerică între două sau mai multe variabile statistice, care, în general, au o dispersie asemănătoare, să fie considerată în acelaşi timp şi legătură de corelaţie, este indispensabilă analiza raporturilor calitative dintre fenomenele cercetate. Numai în urma unei astfel de analize, cunoscîndu-se raporturile obiective dintre fenomene, dependenţa între două sau mai multe caracteristici statistice se poate aprecia ca fiind o legătură de corelaţie, şi, în consecinţă, se poate trece la aplicarea metodelor de interpretare şi măsurare a ei. Pentru a putea reflecta într-o formă cît mai veridică raporturile de cauzalitate existente, influenţele factorilor esenţiali asupra variaţiei caracteristicii rezultative este necesar ca metodele şi procedeele de calcul şi interpretare statistică să se aplice în mod diferenţiat în raport cu posibilităţi de identificare a factorilor-cauză, de cuantificare a fenomenelor, de cuprindere a unităţilor care formează colectivitatea supusă cercetării.

9.2.1. Procedeul seriilor statistice interdependente

Deşi relativ simplu, acest procedeu de analiză a legăturilor dintre fenomene prezintă o serie de avantaje şi dă posibilitatea aplicării în continuare a unor procedee analitice de calcul statistic.

Un prim avantaj pe care-l prezintă acest procedeu de analiză a corelaţiei dintre fenomene este faptul că se pot folosi date din diferite publicaţii, constituite sub formă de serii statistice.

Se pot supune corelaţiei serii statistice de timp, de spaţiu sau de distribuţie.

Acest procedeu de analiză a corelaţiei constă în înscrierea în paralel a unor serii de date statistice, în ordinea raporturilor de dependenţă dintre

Scanned with OKEN Scanner

ele, respectiv se înscriu mai întîi datele cu privire la valorile caracteristicii factoriale (x) și apoi cele ale caracteristicii rezultative (y).

Și aici, ca în orice proces de analiză a corelației, este necesar să se stabilească care este raportul de dependență, respectiv, care este fenomenul ce determină (caracteristica factorială x) și fenomenul determinat (caracteristica rezultativă y).

Și în cazul seriilor statistice interdependente se pot urmări raporturi de corelație simplă sau multiplă.

Pentru folosirea corelației multiple, deci, a verificării interdependenței dintre mai multe variabile statistice, o problemă esențială este aceea de a stabili ordinea de influență a caracteristicilor factoriale. În această situație există două posibilități :

— caracteristicile factoriale sînt independente între ele și atunci ordinea de fixare a lor în analiza raporturilor de interdependență este în funcție de importanța fiecăreia dintre ele pentru procesul de formare a caracteristicii rezultative ;

— caracteristicile factoriale sînt corelate între ele și, în acest caz este necesar ca ordinea de stabilire a factorilor să se facă în raport cu natura obiectivă a existenței legăturii dintre factori.

Procedînd în acest fel se pot analiza nu numai legături directe, imediate, ci și interdependența dintre factori.

Cînd în analiza de corelație se folosesc serii cronologice sau teritoriale, timpul și, respectiv, spațiul nu sînt folosite decît pentru ordonarea seriilor, respectiv pentru obținerea unor serii condiționate în funcție de timp sau de unitatea teritorială, fără ca acestea să fie cuprinse în calculele de corelație.

Folosind acest procedeu se compară vizual dacă există o anumită tendință de corelare a valorilor înscrise pe același rînd. Deci, pe baza acestui procedeu se poate face numai verificarea existenței corelației și a direcției în care ea se manifestă. Pe baza aplicării sale nu se poate verifica nici forma și nici gradul de dependență a variabilei rezultative de factorii care care o determină.

Pentru a măsura intensitatea dependenței, pornind de la seriile interdependente, este necesar să se aplice în continuare metodele analitice de calcul statistic.

9.2.2. Metoda grupărilor

Metoda grupărilor, folosită ca metodă de analiză a legăturilor dintre fenomene, presupune alegerea unei caracteristici factoriale după care să se facă gruparea unităților colectivității și centralizarea valorilor caracteristicilor cu care se corelează. Cu ajutorul acestei metode se pot analiza corelații simple și multiple.

În cazul corelației simple gruparea se face după variația caracteristicii factoriale în funcție de gradul de amplitudine a variației. Caracteristica rezultativă poate să apară cu valori centralizate pe grupele caracteristicii factoriale sau sub formă de distribuții condiționate.

Scanned with OKEN Scanner

De exemplu, dacă se analizează, pe baza grupării, legătura dintre nivelul productivității muncii (x) și retribuția obținută (y), atunci se vor grupa unitățile colectivității în funcție de amplitudinea variației caracteristicii factoriale — productivitatea muncii.

Caracteristica rezultativă va fi prezentată sub formă de medie condiționată, calculată ca raport între suma retribuțiilor din fiecare grupă și frecvența stabilită în funcție de variabila x:

$$\bar{y}_x = \frac{\Sigma y f_{xy}}{\Sigma f_{xy}} \tag{9.2}$$

După modul în care se distribuie mediile condiționate ale variabilei rezultative față de variația caracteristicii factoriale se poate aprecia dacă fenomenele cercetate sînt sau nu corelate și care este direcția în care se manifestă legătura.

În cazul grupării combinate, folosită la analiza corelației simple, se obține un tabel de corelație, care dată fiind importanța folosirii lui în calcule de corelație, se tratează ca un procedeu separat.

În cazul corelației multiple înseamnă că se analizează o singură caracteristică rezultativă variabilă de mai mulți factori. Aici se pot folosi fie grupări simple obținute după variația unei singure caracteristici factoriale, fie grupări combinate după variația mai multor caracteristici factoriale. În prima variantă — gruparea simplă — valorile celorlalte caracteristici apar ca medii condiționate de primul factor de grupare; în cea de-a doua variantă — gruparea combinată — se corelează caracteristicile factoriale în ordinea lor de interdependență și pentru caracteristica rezultativă se calculează numai valorile mediilor condiționate de variația factorilor respectivi de la o grupă la alta. De exemplu, pentru a studia corelația dintre vechimea în producție, îndeplinirea normelor de producție și retribuția obținută, s-a organizat o observare selectivă pentru un eșantion de 200 muncitori.

În acest caz se consideră vechimea în producție ca fiind prima caracteristică factorială (x_1) care este legată în mod obiectiv de cea de-a doua caracteristică factorială — îndeplinirea normelor de producție (x_2), iar retribuțiile obținute (y) pot fi considerate ca dependente de ambii factori.

Dacă se folosește o grupare simplă, atunci cei 200 de muncitori se vor grupa după vechime în grupe de cîte 5 ani — urmînd să se coreleze cu mediile condiționate de variație lui x_1 pentru gradul de îndeplinire a normelor și retribuția medie, adică:

$$\bar{x}_{2/x_1} = \frac{\Sigma x_2 f_{x_1 x_2}}{\Sigma f_{x_1 x_2}} \tag{9.3}$$

și

$$\bar{y}_{/x_1} = \frac{\Sigma y f_{x_1 y}}{\Sigma f_{x_1 y}}. \tag{9.4}$$

În cazul grupării combinate după ambele caracteristici factoriale, fiecare grupă formată după vechime se reîmparte după variația gradului de

Scanned with OKEN Scanner

îndeplinire a normelor, iar pentru retribuţiile obţinute (y) se calculează medii condiţionate de ambii factori, adică :

$$\bar{y}_{/x_1,\, x_2} = \frac{\Sigma y\, f_{x_1 x_2 y}}{\Sigma f_{x_1 x_2 y}}. \qquad (9.5)$$

Dacă pe măsură ce creşte vechimea creşte şi gradul de îndeplinire a normelor şi retribuţia medie, înseamnă că între cele trei variabile există o corelaţie statistică directă, confirmînd faptul că retribuţiile muncitorilor depind şi de vechimea în producţie şi de proporţia în care îşi îndeplinesc normele de producţie.

Metoda grupării are însă şi ea un grad limitat de analiză a corelaţiei. Această metodă dă posibilitatea sistematizării materialului obţinut din observare, în vederea aplicării calculelor de corelaţie. Ca metodă independentă, ea nu permite decît să se constate dacă între caracteristicile studiate există sau nu o corelaţie statistică şi dacă există să se aprecieze direcţia în care se produce, eventual să se sesizeze forma în care se manifestă. Identificarea formei de legătură nu se poate face fără a se întocmi graficul de corelaţie, iar verificarea ei fără efectuarea analizei dispersionale.

Pentru ca metoda grupării să permită aplicarea corectă a calculelor de corelaţie este necesar să se ţină seama de anumite principii, care asigură o grupare ştiinţifică a datelor.

În primul rînd, trebuie să se aleagă pentru efectuarea grupării numai acele caracteristici care corespund unui raport obiectiv de interdependenţă şi pentru care se pot stabili grupele şi intervalele de grupare corespunzătoare scopului analizei. Odată stabilită caracteristica factorială şi cea rezultativă este necesar să se stabilească, în continuare, numărul de grupe şi mărimea intervalelor în funcţie de amplitudinea variaţiei şi de tendinţa manifestată în distribuţia frecvenţelor. Pentru a reda cît mai fidel forma de variaţie a caracteristicilor este indicat să se folosească un număr mai mare de grupe decît în cazul folosirii grupării ca metodă de analiză a structurii colectivităţii.

În al doilea rînd, se recomandă ca, pe cît posibil, să se evite gruparea pe intervale neegale pentru a nu influenţa asupra densităţii de repartiţie a frecvenţelor pe grupe, pentru a nu influenţa în mod subiectiv asupra formei în care se manifestă legătura dintre caracteristicile studiate. În cazul în care sistematizarea datelor se face după două caracteristici de grupare şi se analizează o corelaţie simplă este indicat ca şi pentru caracteristica rezultativă să se folosească tot intervale egale de grupare şi acelaşi număr de grupe, deoarece în calculele de corelaţie apar totdeauna valori perechi, înregistrate la nivelul fiecărei unităţi sau al aceleiaşi grupe.

În al treilea rînd, metoda grupării trebuie să fie combinată cu calcularea mediilor şi a indicatorilor relativi pe grupe. Folosind mediile de grupă se pot înlătura influenţele factorilor neesenţiali, păstrîndu-se ceea ce este esenţial şi tipic în formarea fenomenului studiat, legătura dintre fenomene realizîndu-se, după cum se ştie, numai ca o tendinţă generală, valabilă în medie pentru întregul ansamblu de unităţi statistice.

Scanned with OKEN Scanner

De reţinut că metoda grupării nu trebuie folosită decît în cazul unui număr mare de observaţii statistice, cînd aplicarea metodelor analitice de calcul nu se poate face fără o grupare prealabilă a datelor individuale.

9.2.3. Metoda grafică

Cu ajutorul acestei metode se poate verifica dacă între fenomenele supuse corelaţiei există legături statistice şi dacă ele există se pot aprecia forma şi sensul în care ele se realizează. În plus, metoda grafică se poate folosi şi ca instrument de prezentare a rezultatelor analizei de corelaţie, după ce au fost calculaţi indicatorii care exprimă numeric gradul de legătură dintre fenomenele cercetate, servind astfel pentru compararea datelor distribuţiilor empirice cu valorile teoretice determinate prin aplicarea metodelor analitice de calcul.

Cel mai frecvent, metoda grafică se foloseşte pentru interpretarea legăturilor dintre două fenomene ce se pot exprima sub forma a două variabile statistice. Pentru aceasta se foloseşte sistemul axelor rectangulare, respectiv cadranul I, în care valorile lui x şi y sînt pozitive. Pentru analiza datelor de corelaţie se reprezintă grafic pe axa OX (axa absciselor) — valorile caracteristicii factoriale, iar pe axa OY (axa ordonatelor) — valorile caracteristicii rezultative. Pentru a uşura interpretarea datelor, cele două scări de reprezentare nu încep cu punctul de origine (unde x şi y au valoarea zero) ci, în mod convenţional, cu valorile cele mai apropiate de limitele inferioare înregistrate pentru cele două caracteristici.

La stabilirea scărilor de reprezentare pe cele două axe se recomandă să se asigure o anumită proporţionalitate între ele în raport cu gradul de variaţie a ambelor caracteristici. Dacă se asigură o proporţie justă între cele două scări de reprezentare, atunci graficul este întocmit corect şi, într-adevăr, cu ajutorul lui se va putea prezenta forma obiectivă în care se produce legătura. Stabilind astfel cele două scări de reprezentare se va asigura un cîmp de corelaţie corespunzător tipului de dependenţă dintre cele două variabile. De exemplu, dacă într-un şir de observaţii statistice se constată că există tendinţa ca la o creştere a productivităţii muncii pînă la 100 produse să-i corespundă o creştere a retribuţiei medii pînă la 200 lei, atunci este indicat ca scara de reprezentare a caracteristicii factoriale să fie stabilită din 100 în 100 de unităţi, iar scara de reprezentare a caracteristicii rezultative din 200 în 200 unităţi de lungime.

După ce s-a stabilit scara de reprezentare a celor două variabile se va trece la reprezentarea pe grafic a fiecărei unităţi purtătoare a celor două caracteristici, printr-un punct care va marca intersecţia între valoarea abscisei corespunzătoare caracteristicii factoriale x şi valoarea ordonatei sale, proporţională cu valoarea caracteristicii rezultative y. Dacă se dispune de un număr mare de observaţii statistice, atunci se va obţine pe grafic un număr mare de puncte de intersecţie ale valorilor caracteristicii factoriale x cu cele ale caracteristicii rezultative y. Din această cauză, în literatura de specialitate, acest grafic este denumit „graficul norilor de puncte".

Scanned with OKEN Scanner

Cu ajutorul „graficul norilor de puncte" se pot rezolva, în principal, două aspecte ale analizei corelației dintre fenomene :

a) existența și direcția legăturii ;

b) forma de legătură dintre cele două variabile.

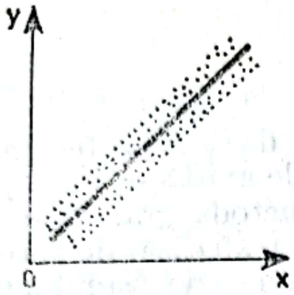


Fig. 9.1. Legătură liniară directă.

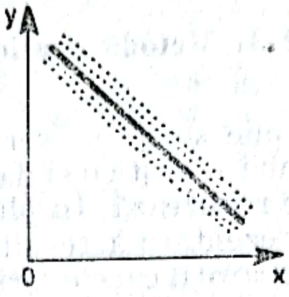


Fig. 9.2. Legătură liniară inversă.

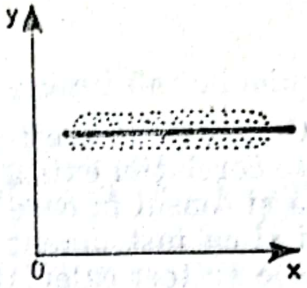


Fig. 9.3. Lipsă de legătură.

Interpretînd modul de împrăștiere a punctelor pe grafic se poate aprecia care este tendința de asociere a celor două caracteristici pentru fiecare unitate în parte. Fenomenele de masă, după cum s-a mai arătat, sînt influențate și de existența unor factori întîmplători, care fac ca regularitățile de producere a legăturilor să se abată uneori de la forma teoretică de manifestare a legăturii. Astfel, analizînd tendința pe care o prezintă pe grafic poziția punctelor corespunzătoare frecvenței de apariție a valorilor individuale ale celor două caracteristici se poate aprecia dacă în medie există legătură, sau fenomenele sînt independente între ele, iar în cazul în care există legătură — care este sensul producerii ei.

Cel mai des , distribuția punctelor pe grafic apare sub cele trei variante prezentate la fig. 9.1, fig. 9.2 și fig. 9.3.

Cînd punctele reprezentînd distribuția bidimensională tind să se concentreze către diagonala dusă din unghiul drept de sus către unghiul stîng de jos (fig. 9.1), legătura este de sens direct, deoarece pe măsură ce cresc valorile caracteristicii factoriale x, cresc și valorile caracteristicii rezultative y.

Cînd tendința de concentrare a norilor de puncte este către cea de-a doua diagonală, care unește unghiul stîng de sus cu unghiul drept de jos (fig. 9.2), atunci legătura este de sens invers, adică pe măsură ce crește valoarea caracteristicii x scade valoarea caracteristicii y.

Și într-un caz și în celălalt se poate aprecia gradul de intensitate a legăturii prin tendința de concentrare către linia teoretică trasată vizual, care unește valorile extreme ale punctelor de intersecție ale valorilor lui x cu cele ale lui y.

În sfîrșit, în cel de-al treilea grafic (fig. 9.3) punctele de intersecție dintre valorile lui x cu cele ale lui y se împrăștie pe întregul cîmp de corelație, fără nici o regularitate. În acest caz fenomenele reprezentate grafic sînt independente între ele și linia OY este paralelă la axa OX.

Deoarece în cele trei variante prezentate, s-a urmărit, în principal, constatarea existenței și direcției legăturii, pentru ușurința înțelegerii graficului, s-a presupus că între cele două variabile x și y există o legătură de tip liniar. Evident, sînt și cazuri în care între x și y pot exista legă-

Scanned with OKEN Scanner

turi de formă neliniară şi pe grafic se prezintă ca o linie curbă; cel mai frecvent apar în legăturile social-economice, hiperbole, parabole de gradul 2 sau ecuaţii exponenţiale (vezi fig. 9.4, 9.5 şi 9.6).

Faţă de celelalte metode prezentate anterior, metoda grafică prezintă avantajul că, cu ajutorul ei se poate constata nu numai existenţa legăturii

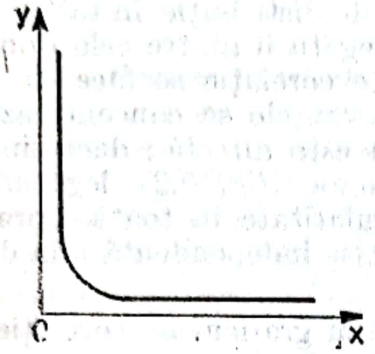


Fig. 9.4. Hiperbolă.

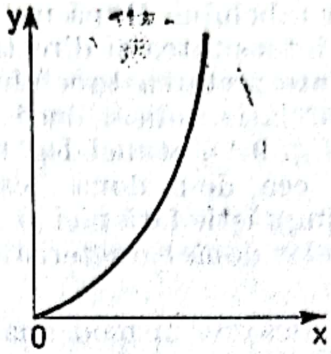


Fig. 9.5. Parabolă.

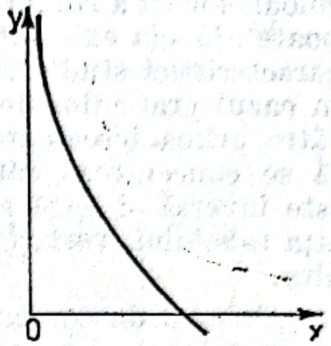


Fig. 9.6. Funcţie exponenţială.

şi sensul ei, dar mai ales forma către care tinde să se realizeze. Metoda grafică se poate folosi, deci, atît ca metodă independentă de prezentare a corelaţiei, cît şi ca instrument de stabilire a formei de legătură, fiind considerată calea cea mai sigură de a alege corect funcţia analitică care se studiază.

9.2.4. Procedeul tabelului de corelaţie

Procedeul tabelului de corelaţie este, de fapt, o combinare a metodei grupării cu principiile de construire şi interpretare a unei reprezentări grafice.

Tabelul de corelaţie este o formă specială de prezentare a rezultatelor unei grupări combinate obţinută după variaţia celor două caracteristici.

În cazul tabelului de corelaţie, gruparea unităţilor se face după ambele caracteristici, atît după variaţia caracteristicii factoriale, cît şi după variaţia caracteristicii rezultative. La întocmirea tabelului de corelaţie trebuie să se respecte anumite reguli stabilite prin analogie cu construcţia graficului de corelaţie ţinînd seama şi de condiţiile impuse de metoda grupării.

În primul rînd, se recomandă ca pentru ambele caracteristici după care se face gruparea să se stabilească intervale egale de grupare, în raport cu amplitudinea variaţiei fiecărei caracteristici.

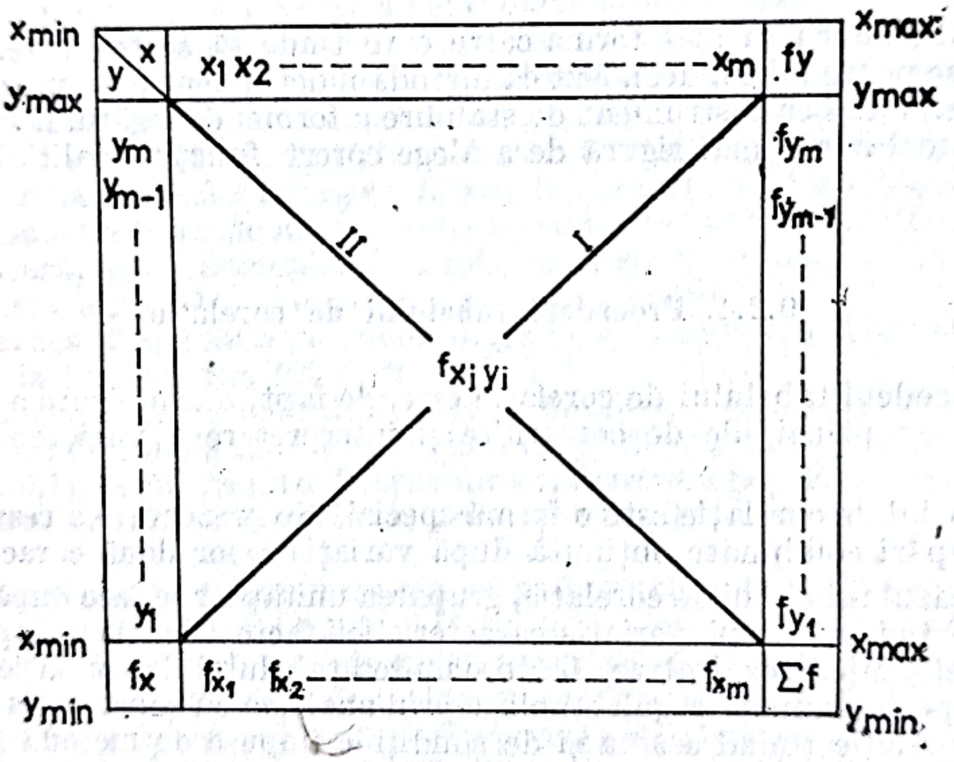
În al doilea rînd, se recomandă să se folosească, pe cît posibil, un număr egal de grupe atît pentru caracteristica factorială, cît şi pentru caracteristica rezultativă. Pentru a desprinde cît mai bine modul de corelare a celor două caracteristici este indicat să se folosească un număr suficient de mare de grupe în scopul asigurării unui cîmp cît mai mare de corelaţie şi al redării corecte a formei obiective a relaţiei de interdependenţă.

În al treilea rînd, se recomandă ca modul de înscriere a celor două caracteristici în tabelul de corelaţie să fie analog modului de prezentare a

Scanned with OKEN Scanner

lor în graficul de corelaţie. Potrivit acestui principiu, grupele după variaţia caracteristicii factoriale se trec pe orizontală, în ordine crescătoare, în capetele coloanelor, iar grupele formate după variaţia caracteristicii rezultative se trec pe verticală, în ordine descrescătoare, în capetele rîndurilor. Numărul unităţilor colectivităţii cercetate, purtătoare a celor două caracteristici (frecvenţele), se trec în rubricile formate din întretăierea coloanelor cu a rîndurilor tabelului. După modul de distribuţie în tabel se poate aprecia existenţa, intensitatea şi direcţia legăturii dintre cele două caracteristici studiate. Interpretarea tabelului de corelaţie se face ca şi în cazul graficului de corelaţie, adică, dacă frecvenţele se concentrează către prima bisectoare (fig. 9.1), atunci legătura este directă; dacă tind să se concentreze către cea de-a doua bisectoare (fig. 9.2) legătura este inversă şi dacă se împrăştie fără nici o regularitate în toată suprafaţa tabelului, variaţia celor două caracteristici este independentă una de alta.

Tabelul de corelaţie alcătuit în mod analog cu graficul de corelaţie, are următoarea schemă:



Acest tabel de corelaţie se prezintă astfel în cele patru unghiuri ale sale:

a) unghiul din stînga jos: x şi y au valori minime;

b) unghiul din stînga sus: x are valoare minimă, iar y are valoare maximă;

c) unghiul din dreapta sus: x şi y au valori maxime;

d) unghiul din dreapta jos: x are valoare maximă, iar y are valoare minimă;

Prima bisectoare (I) marchează, legătura funcţională directă; a doua bisectoare (II) marchează legătura funcţională inversă.

După modul de asociere a nivelurilor celor două caracteristici se distribuie frecvenţele în cadrul tabelului şi se poate interpreta care este tendinţa de formare a raportului de corelaţie între cele două variabile x şi y.

Pentru exemplificare se vor folosi datele publicate de Direcţia Centrală de Statistică cu privire la corelaţia dintre valoarea fondurilor fixe şi numărul de muncitori din întreprinderile industriale în anul 1975 (vezi tabelul 9.1).

Tabelul 9.1

Gruparea întreprinderilor industriale după mărimea fondurilor fixe şi numărul muncitorilor, în anul 1975

Fondurile fixe în mii lei valoare iniţială / Numărul muncitorilor	sub 5 000	5 001 – 20 000	20 001 – 60 000	60 001 – 100 000	100 001 – 200 000	200 001 – 100 000	peste 500 000	Total
A	1	2	3	4	5	6	7	85
peste 5 000 (6 000)	—	—	—	—	1	18	41	60
3 001 – 5 000 (4 000)	—	—	2	4	17	57	39	119
2 001 – 3 000 (2 500)	—	4	14	27	52	55	30	182
1 001 – 2 000 (1 500)	4	38	70	75	105	89	43	424
500 – 1 000 (750)	15	99	91	75	82	33	20	415
201 – 500 (350)	52	109	91	53	41	4	7	357
sub 200 (100)	23	60	75	13	2	1	—	174
Total	94	310	343	247	300	257	180	1 731

Sursa: Prelucrat după Anuarul Statistic al R.S. România, D.C.S., 1976, p. 86–87.

În acest tabel apare clar legătura directă dintre mărimea fondurilor fixe şi numărul de muncitori, deoarece majoritatea frecvenţelor se plasează pe prima diagonală.

Aceeaşi tendinţă de legătură se poate urmări şi pe baza comparaţiei dintre cele două distribuţii marginale. Astfel şi pentru fondurile fixe şi pentru numărul de muncitori ocupaţi într-o întreprindere se manifestă aceeaşi tendinţă de distribuţie a frecvenţelor, adică ponderile cele mai mari le au valorile centrale ale celor două variabile, iar valorile extreme au ponderile cele mai mici.

Faptul că nu există o simetrie perfectă a celor două distribuţii marginale şi nici a celor condiţionate de factorul înregistrat, înseamnă că mărimea fondurilor fixe nu este singurul factor care determină mărimea numărului de muncitori.

Pentru a sesiza mai bine semnificaţia şi forma de legătură trebuie să se calculeze în continuare şi mediile condiţionate, obţinîndu-se o repartiţie bidimensională de frecvenţe.

Scanned with OKEN Scanner

Pe baza mediilor condiționate s-a obținut o distribuție bidimensională cu frecvențe comune, care arată clar că pe măsură ce crește mărimea fondurilor fixe crește și numărul mediu al muncitorilor ce revin pe o întreprindere din fiecare grupă formată în funcție de factorul înregistrat.

Tabelul 9.2

Corelația dintre mărimea fondurilor fixe și mărimea numărului de muncitori în industria R.S. România, în anul 1975

Grupe după mărimea fondurilor fixe (mii lei)	Nr. întreprinderilor	Numărul mediu al muncitorilor, pe grupe	În % față de prima medie ($x_1 = x_0$)
(x_i)	($f_{x_i y_i}$)	($\bar{y}/x_i$)	$\bar{y}\,x_i/(\bar{y}x_0)$
1	2	3	4
sub 5 000	94	402	100
5 001— 20 000	310	598	148,8
20 001— 60 000	343	745	185,3
50 001—100 000	247	1 102	274,1
100 001—200 000	300	1 459	362,9
200 001—500 000	257	2 456	610,9
peste 500 000	180	3 105	772,4
Total	1 731	1 374	—

Pentru a ilustra mai bine dependența statistică dintre cele două variabile se vor prezenta grafic și datele din tabelul de corelație și distribuția mediilor condiționate ale numărului de muncitori în funcție de mărimea fondurilor fixe (vezi fig. 9.7).

Se poate spune că tabelul de corelație este un instrument principal de interpretare a legăturii dintre fenomene, fără a putea însă ca pe baza lui să se determine cu certitudine care este forma de legătură și nici gradul de intensitate în care ea se realizează. Pentru precizarea formei de legătură, datele din tabelul de corelație sînt reprezentate grafic, iar pentru stabilirea gradului de intensitate a corelației se aplică metodele analitice de calcul statistic. Fiind intervale negale, se vor prezenta grafic punctele de coordonate dintre centrul de interval (x) și media condiționată y.

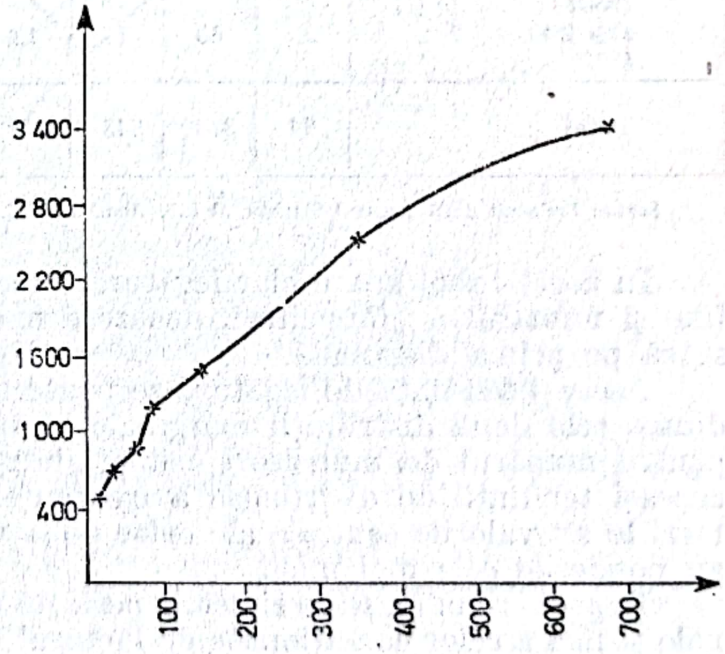


Fig. 9.7. Corelația dintre mărimea fondurilor fixe și numărul muncitorilor în industria R.S. România, în anul 1975.

9.2.5. Metoda balanțelor

Spre deosebire de metodele prezentate anterior, care se referă la analiza interdependenței dintre variabile statistice, metoda balanțelor servește pentru analiza relațiilor care există în cadrul unui proces stohastic. În cazul unui proces stohastic, se pot analiza relațiile de interdependență dintre diferitele elemente ale procesului, dintre diferitele laturi ale lui sau dintre diferitele etape sau momente în care el se desfășoară.

Ca metodă, balanțele sînt, de fapt, o dezvoltare a metodei grupărilor aplicate, nu unor date individuale, ci unor mărimi sintetice, caracteristice procesului social-economic analizat. Și în acest caz metoda grupării pe elemente sau laturi ale procesului se folosește împreună cu mărimile medii și mărimile relative, care permit reliefarea legăturilor obiective existente ce se formează la nivelul întregului ansamblu, prin compensarea abaterilor individuale din interiorul procesului analizat.

În forma sa cea mai generală, metoda balanței este folosită pentru analiza procesului reproducției socialiste lărgite a întregii economii. Cu ajutorul ei se interpretează corelațiile care se formează în mod obiectiv între elementele procesului de reproducție, pe ramuri și forme de proprietate; corelațiile dintre producție, consum și acumulare; între reproducția avuției naționale și a produsului social; între venitul național și nivelul de trai etc.

Un loc din ce în ce mai important, în sistemul de balanțe al economiei naționale, îl ocupă balanța legăturilor dintre ramuri — întocmită sub formă de balanță șah.

În cadrul balanței șah, aceleași grupe se trec atît în subiectul, cît și în predicatul tabelului, iar în rubricile ei se înregistrează schimbarea unui fenomen care are ca puncte de plecare grupele din predicatul tabelului. Aceasta înseamnă că în subiectul tabelului se vor întîlni ramurile în calitatea lor de producători, iar în predicatul tabelului aceleași ramuri în calitate de consumatori sau de elemente de acumulare. Pe baza datelor din balanță se pot caracteriza multiplele legături reciproce ce se formează în mod obiectiv pe linie de producție — consum — acumulare, care iau naștere între grupele colectivității în cadrul procesului reproducției socialiste. Balanța șah permite, deci, caracterizarea în expresie numerică a legăturilor de corelație multiplă ce agreghează într-un tot unitar întregul mecanism al ramurilor și subramurilor, pe unități administrative și economice ale economiei unei țări. Caracterizarea acestor interdependențe, deosebit de complexe, bazate pe un bogat material informativ, se face cu ajutorul calculului matriceal, folosind mijloacele moderne de calcul.

Metoda balanțelor este, deci, caracteristică analizei proceselor sociale și economice ce se produc pe baza unor legături de tip stohastic, a căror interpretare nu este posibilă decît prin folosirea unui sistem de indicatori sintetici, stabiliți pe fiecare grupă și pe întreaga colectivitate, ca mărimi totalizatoare sau medii. Prin interpretarea datelor din balanță se pot stabili legăturile, precum și gradul de intensitate a acestora, prin utilizarea unor indicatori de corelație care să exprime numeric relațiile ce se manifestă în interiorul procesului social-economic analizat.

Scanned with OKEN Scanner

9.3. METODE ANALITICE DE MĂSURARE ŞI ANALIZĂ A LEGĂTURILOR STATISTICE

În paragraful precedent au fost prezentate metodele de verificare a legăturilor dintre fenomenele şi procesele economice de masă, arătîndu-se cu această ocazie care sînt cazurile în care se pot aplica, precum şi limitele în care se pot formula unele concluzii cu privire la interpretarea conexiunii dintre fenomene. Aceste metode sînt folosite, în special, ca momente iniţiale ale cercetării statistice a corelaţiilor dintre fenomene; cu ajutorul lor se face o sistematizare a informaţiilor necesare aplicării metodelor analitice de calcul şi interpretare statistică.

Aplicarea metodelor analitice trebuie să aibă în vedere, în primul rînd, aspectele scoase în evidenţă în urma efectuării analizei calitative care trebuie să preceadă orice fel de cercetare statistică, deci, cunoaşterea conţinutului social-economic al relaţiilor de cauzalitate dintre fenomene şi, în al doilea rînd, să folosească cu discernămînt concluziile formulate prin aplicarea metodelor simple de corelaţie, prin care s-au verificat existenţa, direcţia şi forma în care se realizează legătura dintre fenomene.

În cazul în care nu se ţine cont de aceste principii, calculul statistic devine un simplu exerciţiu formal pentru determinarea unor indicatori de corelaţie nesemnificativi.

Posibilităţile de aplicare a metodelor analitice de corelaţie, depind, deci, de natura specifică a fenomenelor cercetate, de volumul datelor de care se dispune, respectiv, de seriile de distribuţie formate prin centralizarea datelor şi de numărul caracteristicilor luate în studiu.

În cazul unei serii de distribuţie bidimensionale, cînd există o variabilă factorială (x) şi o variabilă rezultativă (y), se pune problema să se calculeze indicatori pentru o corelaţie simplă. În cazul cînd există o distribuţie multidimensională, cu două, trei sau mai multe variabile factoriale ($x_1, x_2, \ldots, x_n$) şi o variabilă rezultativă (y), trebuie studiată prin metode statistice o corelaţie multiplă. În continuare se vor prezenta, pe rînd, cele două situaţii.

9.3.1. Corelaţia liniară simplă

În cazul corelaţiei simple este necesar să se urmărească legătura dintre o caracteristică statistică considerată ca variabilă independentă (x) şi o altă caracteristică statistică considerată drept variabilă dependentă sau rezultativă (y). Într-o legătură statistică, caracteristica rezultativă fiind dependentă de mai mulţi factori este necesar să se aleagă pentru a fi studiată interdependenţa care se realizează între o singură caracteristică factorială şi cea rezultativă. Această caracteristică factorială trebuie să aibă însă un caracter esenţial în procesul de formare a caracteristicii rezultative, pentru a putea obţine indicatori de corelaţie cu valori ridicate. Fiind vorba de fenomenele de masă şi legăturile de corelaţie se exprimă tot sub formă de tendinţă medie valabilă pentru un număr mare de cazuri individuale diferite, dar aparţinînd aceleaşi colectivităţi.

Scanned with OKEN Scanner

Pentru a evidenția legea care se manifestă în fiecare legătură în parte, pentru a măsura statistic tendința sa de manifestare se folosesc ecuațiile de estimare corespunzătoare unei funcții analitice care exprimă forma de legătură dintre caracteristica factorială și cea rezultativă. Această funcție este cunoscută sub denumirea de funcție de regresie, iar reprezentarea ei grafică se face prin linia (curba) de regresie. Alegerea corectă a funcției de regresie, care să exprime cel mai bine legătura dintre cele două caracteristici, este deosebit de importantă pentru determinarea valorii statistice a indicatorilor de corelație. Forma legăturii dintre cele două caracteristici se exprimă printr-o ecuație de tendință în care valorile caracteristicii rezultative se calculează în funcție de valorile caracteristicii factoriale aleasă. Dacă funcția de regresie nu a fost aleasă în mod corespunzător, atunci rezultatele analizei, respectiv, valorile indicatorilor de corelație vor fi denaturate.

Pentru alegerea corectă a funcției de regresie este necesar să se reprezinte cele două serii de distribuție printr-un grafic de corelație, care să permită interpretarea vizuală a tendinței de asociere a celor două caracteristici perechi. Pe baza reprezentării grafice se poate aprecia apoi dacă legătura este de formă rectilinie sau curbilinie. Tot ca instrument de verificare a veridicității funcției de estimare, pentru măsurarea legăturii dintre cele două variabile în cadrul colectivității înregistrate, se pot folosi și metodele de analiză dispersională, cu interpretarea corespunzătoare a rezultatelor obținute în urma aplicării ei.

Funcția de regresie exprimă statistic modul în care caracteristica rezultativă (y) s-ar modifica dacă ar varia numai valorile caracteristicii factoriale (x) iar ceilalți factori ar fi considerați cu acțiune constantă în toate cazurile observate. Celelalte caracteristici, fiind considerate ca neesențiale și cu acțiune constantă asupra tuturor unităților la care se măsoară raportul de interdependență, înseamnă că influența lor este sintetizată într-o singură valoare cu caracter de medie. Legătura dintre cele două variabile se realizează tot sub formă de tendință, ceea ce înseamnă că pentru estimare se va folosi tot o ecuație medie de tendință corespunzătoare formei identificate pe grafic.

În cazul în care, prin reprezentarea grafică, se observă o tendință de legătură de tip liniar, în care variația caracteristicii rezultative prezintă o anumită tendință de uniformitate a modificării sale absolute sub influența caracteristicii considerată ca fiind determinantă, ecuația care exprimă această formă de legătură va fi :

$$\bar{Y}_x = a + bx. \tag{9.6}$$

Ecuația de regresie ($\bar{Y}_x$) se notează, de obicei, ca medie, deoarece mărimea sa exprimă tendința de realizare a corelației dintre cele două variabile x și y. Dacă, într-adevăr, legătura este liniară și factorul este determinant, atunci valorile ecuațiilor de regresie, calculate pentru toate unitățile observate pe baza valorii individuale ale variabilei x, trebuie să prezinte

Scanned with OKEN Scanner

abateri minime față de valorile empirice. Pentru măsurarea tendinței de realizare a legăturii, în ecuația mediei de regresie liniară cei doi parametri au și ei conținut de valori medii și trebuie să fie reprezentativi pentru cele mai multe din unitățile observate.

Parametrul a are caracter de mărime medie, în sensul că valoarea sa arată la ce nivel ar fi ajuns valoarea caracteristicii y dacă toți factorii — mai puțin cel înregistrat — ar fi avut o acțiune constantă asupra formării ei. În acest caz valorile individuale ale caracteristicii rezultative ar fi fost egale între ele și, deci, egale cu media lor.

Parametrul b se mai numește și *coeficient de regresie* și exprimă, în sens geometric, panta liniei drepte. Coeficientul de regresie b arată care este gradul de influență a caracteristicii alească drept caracteristică factorială x și măsoară cu cît se schimbă în medie variabila y în cazul în care variabila x crește cu o unitate.

Coeficientul de regresie arată nu numai gradul de influență a factorului x asupra variabilei y, ci și sensul în care se realizează legătura. Valoarea coeficientului de regresie b poate fi mai mare decît zero, mai mică decît zero și egală cu zero.

În cazul în care $b > 0$, atunci legătura de corelație este directă, deoarece pe măsură ce cresc valorile lui x, cresc și valorile ecuației de regresie calculate.

În cazul în care $b < 0$, legătura este de sens invers, adică pe măsură ce crește valoarea caracteristicii-factor, scade valoarea caracteristicii rezultative.

Cînd $b = 0$, cele două variabile sînt independente între ele și atunci $\bar{Y}_x = a$, deci, valoarea medie a ecuației de regresie este egală cu valoarea medie a caracteristicii rezultative ($\bar{Y}_x = y$). Aceasta înseamnă că variabila y nu este în funcție de variabila x, care s-a considerat ca factor de influență, ci variația ei depinde de ceilalți factori, care s-au considerat ca fiind cu acțiune constantă pentru toate unitățile colectivității.

Pentru a determina ecuația medie de regresie, și cu ajutorul ei, valorile ecuației individuale de regresie corespunzătoare tuturor valorilor variabilei x, este necesar să se calculeze valorile celor doi parametri a și b.

Dacă factorul x este determinat pentru variabila y, atunci valorile estimate prin funcția de regresie trebuie să dea abateri minime față de cele înregistrate pentru variabila rezultativă. Cum aceste abateri se pot produce într-un sens sau altul, ele sînt ridicate la pătrat și, din această cauză, metoda de verificare a acestei condiții se mai numește și metoda celor mai mici pătrate.

Cu alte cuvinte, dacă y depinde de x, atunci trebuie să se îndeplinească condiția ca suma pătratelor abaterilor valorilor empirice de la valorile ecuațiilor lor de regresie să fie minimă:

$$S = \Sigma(y_i - Y_{x_i})^2 = \text{minim.} \qquad (9.7)$$

Pentru tendința liniară această ecuație este:

$$S = \Sigma(y_i - a - bx_i)^2 = \text{minim.}$$

Pentru aflarea celor doi parametri care definesc ecuația liniei drepte se derivează această sumă în raport cu derivatele celor doi parametri:

$$\frac{\delta_s}{\delta_a} = 2\ \Sigma(a - bx - y)\cdot 1;$$

$$\frac{\delta_s}{\delta_b} = 2\Sigma(a - bx - y)\cdot x.$$

Anulînd derivatele parțiale, simplificînd cu 2, înmulțind cu cei doi coeficienți 1 și x și însumînd ecuațiile pentru toate valorile caracteristicilor x și y, se obține:

$$\Sigma a + \Sigma bx - \Sigma y = 0;$$

$$\Sigma ax + \Sigma bx^2 - \Sigma xy = 0.$$

Scoțînd în factori pe a și b și trecînd termenul liber y în membrul doi, se obține sistemul:

$$\begin{aligned} na + b\Sigma x &= \Sigma y; \\ a\Sigma x + b\Sigma x^2 &= \Sigma xy. \end{aligned} \tag{9.8}$$

Sistemul de ecuații normale necesar rezolvării ecuației funcției de regresie se poate obține cu ușurință dacă se înmulțește ecuația dreptei, pe rînd, cu coeficienții celor doi parametri a și b, și se însumează ecuațiile astfel obținute pentru toate unitățile la care s-a făcut observarea socotindu-se ca factori a și b, adică:

— se înmulțește pe rînd, ecuația cu coeficienții lui a și b $(1; x)$ și se obține:

$$\begin{cases} a + bx = y & \text{prima ecuație;} \\ ax + bx^2 = xy & \text{a doua ecuație;} \end{cases}$$

— se însumează toate ecuațiile corespunzătoare celor n termeni, se dă factor comun pe a și b și se obține sistemul de ecuații normale (7) care permite determinarea valorii celor doi parametri:

$$\begin{cases} na + b\Sigma x = \Sigma y; \\ a\Sigma x + b\Sigma x^2 = \Sigma xy. \end{cases}$$

Acest procedeu mai simplu permite obținerea cu ușurință a sistemului de ecuații normale și pentru corelația liniară multiplă, unde se lucrează cu un număr mai mare de parametri corespunzător numărului caracteristicilor factoriale.

Scanned with OKEN Scanner

Rezolvînd sistemul de ecuaţii se obţine valoarea lui a şi b şi se calculează valoarea ecuaţiei de regresie pentru fiecare valoare a caracteristicii x. Aceste valori ale ecuaţiilor de regresie se mai numesc şi valorile teoretice ale caracteristicii y în funcţie de x, iar operaţia de înlocuire a termenilor reali y cu valorile ecuaţiilor de regresie (valori teoretice) se numeşte *ajustare*.

Deci, prin ajustarea unei serii statistice de distribuţie se înţelege înlocuirea termenilor empirici (termeni reali obţinuţi prin observare) cu termeni teoretici, care arată tendinţa de variaţie a caracteristicii rezultative, dacă ar fi depins numai de variaţia lui x considerat.

În analiza corelaţiei dintre fenomene valorile ecuaţiilor de regresie au un rol important. Ele pot fi considerate ca valori teoretice, care exprimă tendinţa de manifestare a interdependenţei dintre fenomene, dacă în procesul de interacţiune dintre ele nu ar interveni şi alţi factori esenţiali sau întîmplători care să modifice gradul de legătură pentru fiecare unitate în parte.

Pentru exemplificare se va folosi legătura care există între dinamica productivităţii muncii şi dinamica retribuţiei medii a personalului muncitor din industrie, între anii 1970—1976.

Tabelul 9.3

Dinamica productivităţii muncii şi a retribuţiei medii în industrie, în perioada 1970—1976
(1965 = 100)

Indici	1970	1971	1972	1973	1974	1975	1976
A	1	2	3	4	5	6	7
Indicii productivităţii muncii	142	149	157	169	181	194	210
Indicii retribuţiei medii	123	126	128	134	143	157	167

Sursa: Anuarul Statistic al R.S. România, D.C.S. 1975, p. 76 şi 141 şi 1977, p. 110 şi 179.

Pentru a măsura corelaţia dintre cele două variabile este necesar să se determine funcţia analitică care o caracterizează. Pentru a alege corect funcţia de regresie corespunzătoare legăturii dintre cele două variabile s-au reprezentat grafic cele două serii de date statistice. Din graficul de corelaţie se apreciază că legătura este de tip liniar şi că, deci, ecuaţia medie a funcţiei de regresie este :

$$\bar{Y}_x = a + bx.$$

Pentru a afla valorile celor doi parametri a şi b este necesar să se rezolve sistemul de ecuaţii normale. Dacă se foloseşte metoda determinanţilor se obţine :

$$a = \frac{\begin{vmatrix} \Sigma y & \Sigma x \\ \Sigma xy & \Sigma x^2 \end{vmatrix}}{\begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}} = \frac{\Sigma y \ \Sigma x^2 - \Sigma xy \ \Sigma x}{n \ \Sigma x^2 - (\Sigma x)^2} \qquad (9.9)$$

Scanned with OKEN Scanner

$$b = \frac{\begin{vmatrix} n & \Sigma y \\ \Sigma x & \Sigma xy \end{vmatrix}}{\begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}} = \frac{n\,\Sigma xy - \Sigma x\,\Sigma y}{n\,\Sigma x^2 - (\Sigma x)^2}. \qquad (9.10)$$

Urmărind elementele necesare rezolvării acestui sistem de ecuaţii, se vor înscrie datele într-un tabel de calcul (tabelul 9.4).

Tabelul 9.4

Estimarea dinamicii retribuţiei medii din industrie, pe perioada 1970—1976, în funcţie de dinamica productivităţii muncii

Nr. crt.	Dinamica productivităţii muncii (x_i)	Dinamica retribuţiei medii (y_i)	x_i^2	$x_i y_i$	y_i^2	$Y_{x_i} = a + bx_i$ $Y_{x_i} = 24{,}83 + 0{,}67\ x_i$
0	1	2	3	4	5	6
1	142	123	20 164	17 466	15 129	120
2	149	126	22 201	18 774	15 876	124,6
3	157	128	24 649	20 096	16 384	130,0
4	169	134	28 561	22 646	17 956	138,0
5	181	143	32 761	25 883	20 449	146,0
6	194	157	37 636	30 458	24 649	154,7
7	210	167	44 100	35 070	27 889	165,3
	1 202	978	210 072	170 393	138 332	978,6
	Σx_i	Σy_i	Σx_i^2	$\Sigma x_i y_i$	Σy_i^2	$\Sigma Y x_i$

Avînd calculate toate elementele, sistemul de ecuaţii va fi:

$$7\ a + 1\,202\ b = 978;$$

$$1\,202\ a + 210\,072\ b = 170\,393.$$

Rezolvînd sistemul de ecuaţii se obţin rădăcinile: $a = 24{,}83$ şi $b = 0{,}67$.

Deci, ecuaţia medie de estimare a legăturii liniare dintre dinamica productivităţii muncii şi dinamica retribuţiei medii a personalului muncitor din industrie, pe perioada 1970—1976 a fost:

$$\bar{Y}_x = 24{,}83 + 0{,}67\ b.$$

Pe baza ecuaţiei medii de tendinţă se poate spune că, pe perioada luată în calcul, la o creştere cu 1% a nivelului productivităţii muncii, retribuţia medie a crescut în medie faţă de anul precedent cu 0,67 (valoarea coeficientului de regresie b).

Scanned with OKEN Scanner

Pentru verificarea calculului parametrilor funcţiei de regresie se foloseşte relaţia $\Sigma_{y_t} = \Sigma_{y_{x_i}}$ (vezi ultima coloană din tabelul 9.4), ceea ce arată că prin ajustare nu se face altceva decît o redistribuire a influenţei factorilor, astfel încît factorul înregistrat să influenţeze sistematic în toate cazurile supuse observării statistice.

Funcţia de regresie este însă numai o ipoteză statistică care exprimă regularitatea, tendinţa medie de manifestare a legăturii dintre cele două caracteristici, considerînd ca variabil numai factorul înregistrat. În realitate, valorile empirice ale variabilei dependente se formează sub influenţa tuturor factorilor, inclusiv a factorului înregistrat. De aceea, între valorile empirice şi cele estimate pot să apară abateri mai mari sau mai mici, după cum influenţa celorlalţi factori consideraţi cu caracter întîmplător este mai mare sau mai mică.

Deşi valorile ecuaţiilor de regresie permit o apreciere asupra modului de realizare a legăturii dintre cele două caracteristici, prin interpretarea mărimii abaterilor lor, faţă de valorile empirice, totuşi ele nu reprezintă decît un prim pas pentru măsurarea corelaţiei dintre fenomene.

9.3.1.2. RAPORTUL DE CORELAŢIE

Pentru stabilirea intensităţii legăturii dintre cele două caracteristici statistice trebuie să se calculeze un indicator sintetic de corelaţie. Pentru corelaţia liniară simplă se calculează *raportul (indicele) de corelaţie şi coeficientul de corelaţie.*

Raportul de corelaţie este indicatorul care permite măsurarea gradului de intensitate a realizării legăturii dintre caracteristica considerată factor de influenţă şi caracteristica rezultativă. Calculul său se bazează pe descompunerea dispersiei totale a caracteristicii y pe factori de influenţă.

După cum se ştie, dispersia totală a unei caracteristici depinde de gradul de influenţă a tuturor factorilor, deoarece calculul ei se bazează pe distribuţia abaterilor individuale ale tuturor valorilor empirice de la media lor. Deci, valoarea dispersiei totale va cuprinde atît influenţa tuturor factorilor întîmplători, cît şi influenţa celor esenţiali (care au caracter variabil).

Dacă în studiul variaţiei caracteristicii rezultative s-a aplicat regresia liniară simplă, atunci valorile individuale ale ecuaţiilor de regresie vor estima influenţa factorului înregistrat. Valori diferite ale ecuaţiilor de regresie apar numai dacă există influenţa factorului considerat ($b \neq 0$), deoarece ceilalţi factori esenţiali şi întîmplători au fost sintetizaţi în valoarea parametrului a, care intră în mod constant în fiecare ecuaţie de regresie în parte.

În cazul calculării valorilor individuale ale ecuaţiilor de regresie se determină, deci, care ar fi valoarea caracteristicii rezultative dacă ar fi fost influenţată numai de acţiunea factorului x. Rezultă că se pot calcula două şiruri de abateri : abateri între valorile empirice şi ecuaţiile corespunzătoare de regresie ($y_i - Y_x$) şi abateri între ecuaţiile de regresie şi media caracteristicii ($Y_x - \bar{y}$). Deci, dispersia calculată pe baza abaterilor dintre valorile empirice ale caracteristicii y şi valorile ajustate va sintetiza influenţa celorlalţi factori în afara factorului x şi este o parte a dispersiei totale.

Scanned with OKEN Scanner

Dacă se notează dispersia totală a caracteristicii rezultative cu σ_y^2, dispersia față de linia de regresie cu $\sigma_{y/r}^2$ și dispersia liniei de regresie de la valoarea medie a caracteristicii cu $\sigma_{y/x}^2$, atunci între cele trei dispersii există relația:

$$\sigma_y^2 = \sigma_{y/r}^2 + \sigma_{y/x}^2 \tag{9.11}$$

adică:

$$\frac{\Sigma(y_i - \bar{y})^2}{n} = \frac{\Sigma(y_i - Y_{x_i})^2}{n} + \frac{\Sigma(Y_{x_i} - \bar{y})^2}{n}, \tag{9.12}$$

în care:

σ_y^2 arată influența variației tuturor factorilor;

$\sigma_{y/r}^2$ arată influența factorilor considerați cu acțiune constantă și față de influența factorului ales x reprezintă dispersia reziduală;

$\sigma_{y/x}^2$ arată influența factorului (x).

Relațiile dintre dispersii, valabile la nivelul întregului ansamblu, se păstrează și la nivelul unităților, unde abaterea totală individuală se descompune în cele două părți componente:

$$(y_i - \bar{y}) = (y_i - Y_{x_i}) + (Y_{x_i} - \bar{y}). \tag{9.13}$$

În cazul în care între cele două caracteristici x și y există o legătură funcțională, atunci valorile teoretice ar corespunde cu valorile empirice ale caracteristicii y, deoarece acestea s-ar forma numai în funcție de variația caracteristicii x. În acest caz $a = 0$, deci nu există alți factori care influențează în afară de x. Dacă valorile teoretice coincid cu valorile empirice ale caracteristicii y, înseamnă că abaterile valorilor empirice de la valorile ajustate sînt egale cu zero și, deci, dispersia față de valorile de regresie $\sigma_{y/r}^2$ este și ea egală cu zero.

În realitate, în cadrul fenomenelor de masă, legăturile funcționale se întîlnesc foarte rar și numai la nivelul unităților simple. Gradul de intensitate a corelației dintre fenomene se va obține stabilind greutatea specifică a dispersiei formată pe baza factorului înregistrat față de dispersia totală. Acest indicator se numește *coeficientul de determinație*, se notează cu $R_{x,y}^2$ și se calculează după formula:

$$R_{x,y}^2 = \frac{\sigma_{y/x}^2}{\sigma_y^2}. \tag{9.14}$$

În mod analog se calculează și *coeficientul de nedeterminație* (K_{xy}^2) ca raport între dispersia față de linia de regresie și dispersia totală:

$$K_{x,y}^2 = \frac{\sigma_{y/r}^2}{\sigma_y^2}. \tag{9.15}$$

Suma celor doi coeficienți calculați ca mărimi relative de structură față de aceeași bază este egală cu 1:

$$R_{x,y}^2 + K_{x,y}^2 = 1. \tag{9.16}$$

Scanned with OKEN Scanner

Coeficientul de nedeterminație se mai poate calcula și din această relație, adică :

$$R^2_{x,y} = 1 - K^2_{x,y}, \tag{9.17}$$

de unde :

$$R^2_{x,y} = 1 - \frac{\sigma^2_{y/r}}{\sigma^2_y}.$$

Pentru a se aduce la același grad cu valorile empirice și cele teoretice, se extrage rădăcina pătrată din coeficientul de determinație și se obține *raportul (indicele) de corelație* ($R_{x,y}$), care se poate calcula fie pe baza dispersiei formată de factorul determinant ($\sigma^y_{y/x}$), fie pe baza dispersiei calculată față de linia de regresie ($\sigma^2_{y/r}$) :

$$R_{xy} = \sqrt{\frac{\sigma^2_{y/x}}{\sigma^2_y}} = \sqrt{\frac{\frac{\Sigma(Yx_i - \bar{y})^2}{n}}{\frac{\Sigma(y_i - \bar{y})^2}{n}}} = \sqrt{\frac{\Sigma(Yx_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2}}, \tag{9.18}$$

sau :

$$R_{xy} = \sqrt{1 - \frac{\sigma^2_{y/r}}{\sigma^2_y}} = \sqrt{1 - \frac{\frac{\Sigma(y_i - Yx_i)^2}{n}}{\frac{\Sigma(y_i - \bar{y})^2}{n}}} =$$

$$= \sqrt{1 - \frac{\Sigma(y_i - Yx_i)^2}{\Sigma(y_i - \bar{y})^2}}, \tag{9.19}$$

Raportul de corelație poate lua valori de la zero la ± 1. Cînd dispersia termenilor reali de la cei teoretici este zero sau tinde către zero, coeficientul de nedeterminație ($K^2_{x,y}$) este egal cu zero sau tinde către zero, deci, din 1, care indică legătura funcțională, se va scădea o cantitate foarte mică și rădăcina pătrată calculată din această valoare foarte apropiată de unitate va tinde și ea către 1. Dacă cele două caracteristici sînt independente, coeficientul de regresie b din ecuația dreptei va fi egal cu zero și, deci valorile teoretice vor fi egale cu media caracteristicii, deci, $\sigma^2_{y/r} = \sigma^2_y$, iar raportul dintre ele va fi egal cu 1 și deci, $R^2_{x,y} = 0$.

În ceea ce privește semnul raportului de corelație, acesta se va determina după semnul lui b din ecuația de regresie. Dacă b este pozitiv, $R_{x,y}$ va fi luat cu sens pozitiv ; dacă b este negativ, atunci $R_{x,y}$ se va lua cu semnul minus, iar dacă b este egal cu zero trebuie ca și $R_{x,y}$ să fie egal cu zero.

Pentru calcularea raportului de corelație se vor folosi datele din exemplul prezentat anterior, pentru care este necesar să se determine, în prealabil, media caracteristicii y, după formula :

$$\bar{y} = \frac{\Sigma y}{n} = \frac{978}{7} = 139{,}7\ \%.$$

Scanned with OKEN Scanner

Avînd toate datele se va trece la calculul celor două dispersii necesare determinării valorii raportului de corelaţie (tabelul 9.5).

Tabelul 9.5

Calcularea dispersiilor faţă de medie şi faţă de linia de regresie

y_i	Y_{x_i}	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$y_i - Y_{x_i}$	$(y_i - Y_{x_i})^2$
1	2	3	4	5	6
123	120	−16,7	278,89	−3	9,0
126	124,6	−13,7	187,69	1,4	1,96
128	130,0	−11,7	136,89	2	4,0
134	138,0	−5,7	32,49	−4	16,0
143	146,0	3,3	10,89	−3	9,0
157	154,7	17,3	299,29	2,3	5,29
167	165,3	27,3	745,29	1,7	2,89
978	978,6	—	1 691,43	—	39,96
Σy_i	ΣY_{x_i}	—	$\Sigma(y_i - \bar{y})^2$	—	$\Sigma(y_i - Y_{x_i})^2$

Deci :

$$R_{x,y} = \sqrt{1 - \frac{\Sigma(y_i - Y_{x_i})^2}{\Sigma(y_i - \bar{y})^2}} = \sqrt{1 - \frac{39,96}{1\,691,43}} = 0,988.$$

Se poate aprecia că, în cadrul acestui exemplu, legătura de corelaţie este foarte strînsă deoarece raporul de corelaţie are o valoare foarte ridicată (0,988).

În ecuaţia de regresie coeficientul de regresie b a fost pozitiv (+0,67), deci, se va reţine pentru $R_{x,y}$ numai valoarea pozitivă, care arată că între cele două variabile a existat o legătură directă, adică pe măsură ce creşte productivitatea muncii, corespunzător funcţiei de regresie creşte şi retribuţia medie fără ca între cele două variabile să existe o legătură funcţională.

Folosind valoarea raportului de determinaţie ($R^2_{x,y}$) se poate afirma că dintre factorii care au determinat dinamica retribuţiei medii pe perioada 1970—1976, dinamica productivităţii muncii are o contribuţie de 96,74%.

În exemplul prezentat s-a studiat formarea ecuaţiilor de regresie şi a raportului de corelaţie pentru un număr de observaţii la care, pentru fiecare valoare individuală a lui x, a corespuns o singură valoare a caracteristicii y.

În practica statistică, de regulă, se dispune de un număr mare de cazuri la care se face observarea. În această situaţie se iau în discuţie serii de distribuţie de frecvenţe. În astfel de cazuri distribuţia bidimensională se poate prezenta sub două variante :

a) se întîlnesc aceleaşi frecvenţe pentru ambele caracteristici supuse analizei de corelaţie ;

b) fiecare caracteristică are distribuţia sa de frecvenţă.

În primul caz, caracteristica rezultativă este necesar să fie prezentată sub formă de medie ; în cel de-al doilea caz, distribuţia bidimensională este trecută într-un tabel de corelaţie.

Scanned with OKEN Scanner

În cazul distribuției bidimensionale pentru un număr mare de unități observate, sistemul de ecuații normale, care satisface condiția de minim, este :

$$a\Sigma f_{xy} + b\,\Sigma\, xf_x = \Sigma\, y\, f_y \qquad (9.20)$$

$$a\Sigma\, xf_x + b\Sigma\, x^2 f_x = \Sigma\, yxf_{xy}$$

Pentru primul caz — analiza corelației pentru o serie bidimensională cu frecvențe comune — se pot folosi datele cu privire la interpretarea legăturii dintre volumul mediu de lucrări ce revin unui tractor (ha. a.n.) și costul pe un ha.a.n. în întreprinderile pentru mecanizarea agriculturii.

Tabelul 9.6

Distribuția S.M.A. după volumul mediu de lucrări ce revin unui tractor în ha.a.n. și costurile pe un ha.a.n.

(date convenționale)

Volumul mediu de lucrări ce revin unui tractor ha.a.n. (x)	Costurile pe un ha.a.n. în lei (y)	Numărul unităților f_{xy}
604	60,78	8
558	56,86	3
528	63,29	15
603	59,49	22
534	62,84	37
482	63,44	22
427	64,65	40
372	65,97	112

Avînd toate elementele calculate (tab. 9.7) se obține sistemul de ecuații normale :

$$259\, a + 116\,798\, b = 16\,614,$$

$$116\,798\, a + 54\,487\,846\, b = 7\,447\,700,$$

de unde :

$$a = 74{,}96969 \approx 75$$

$$b = -\,0{,}02449 \approx 0{,}024.$$

Ecuația medie de regresie care exprimă legătura dintre volumul mediu al lucrărilor și costurile pe un ha.a.n., în întreprinderile pentru mecanizarea agriculturii este :

$$\bar{Y}_x = 75 - 0{,}024x,$$

ceea ce înseamnă că, în medie, la fiecare creștere de un ha.a.n., costurile scad în medie pentru aceeași unitate cu 0,024.

Scanned with OKEN Scanner

Tabelul 9.7

Corelaţia dintre volumul de lucrări şi costul mediu pe ha.a.n.

x	y	f	x^2	xy	xf	yf	x^2f	xyf	Y_x	Y_{xf}
1	2	3	4	5	6	7	8	9	10	11
372	66,0	112	138 384	24 552	41 664	7 392	15 499 008	2 749 824	66,1	740,3
427	64,7	40	182 329	27 627	17 080	2 588	7 293 160	1 105 076	64,8	2 592,0
482	63,4	22	232 324	30 559	10 604	1 394,8	5 111 128	672 294	63,4	1 394,8
528	63,3	15	278 784	33 422	7 920	949,5	4 181 760	501 336	62,3	934,5
534	62,8	37	285 156	33 534	19 758	2 323,6	10 550 772	1 240 802	62,2	2 301,4
558	56,9	3	311 364	31 750	1 674	170,7	934 092	952 501	61,6	184,8
603	59,5	22	363 609	35 879	13 266	1 309,0	7 999 398	789 327	60,5	1 331,0
604	60,8	8	364 816	36 723	4 832	486,4	2 918 528	293 786	60,5	483,2
—	—	259	—	—	116 798	16 614	54 487 849	7 447 700	—	16 625
		Σf			Σxf	Σyf	Σx^2f	Σxyf		ΣY_{xf}

Scanned with OKEN Scanner

În continuare, pentru a măsura gradul de intensitate a legăturii dintre cele două variabile se va calcula raportul de corelaţie a distribuţiei de frecvenţe bidimensională :

$$R_{x,y} = \sqrt{1 - \frac{\Sigma (y_i - Y_{xi})^2 f_i}{\Sigma (y_i - \bar{y})^2 f_i}}. \qquad (9.21)$$

Pentru rezolvare se vor înscrie datele necesare într-un tabel de calcul.

Tabelul 9.8

Calculul dispersiei valorilor empirice de la valorile ajustate şi de la valoarea lor medie

y_i	Y_{x_i}	f_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(y_i - \bar{y})^2 f_i$	$(y_i - Y_{x_i})$	$(y_i - Y_{x_i})^2$	$(y_i - Y_{x_i})^2 f_i$
1	2	3	4	5	6	7	8	9
66,0	66,1	112	1,9	3,61	404,32	−0,1	0,01	1,12
64,7	64,8	40	0,6	0,36	14,40	−0,1	0,01	0,40
63,4	63,4	22	−0,7	0,49	10,78	0	0	0
63,3	62,3	15	−0,8	0,64	9,60	1,0	1,0	15,0
62,8	62,2	37	−1,3	1,69	62,53	0,6	0,36	13,32
56,9	61,6	3	−7,2	51,84	155,52	4,7	22,09	66,27
59,5	60,5	22	−4,6	21,16	465,52	1,0	1,00	22,00
60,8	60,5	8	−3,3	10,89	87,12	0,3	0,09	0,72
—	—	259 Σf_i	—	—	1 209,79 $\Sigma(\bar{y} - \bar{y})^2 f$	—	—	118,83 $\Sigma(y_i - Y_x)^2 f_i$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{16\,614}{259} = 64{,}1\,;$$

$$R_{x,y} = \sqrt{1 - \frac{118{,}83}{1\,209{,}79}} = -0{,}95.$$

După cum se poate observa, între cele două variabile a existat o legătură foarte strînsă şi de sens invers. S-a reţinut pentru raportul de corelaţie semnul minus, după semnul coeficientului de regresie.

Repartiţiile bidimensionale pot fi prezentate şi sub forma unui tabel de corelaţie în care apar distribuţii condiţionate după ambele caracteristici.

Exemplificarea se va face folosind datele din tabelul de corelaţie (9.9) cu privire la productivitatea muncii pe un muncitor din activitatea de bază (x) şi costul pe un ha.a.n. (y).

Pentru rezolvare este necesar să se prelungească tabelul de corelaţie făcînd produsele de frecvenţă necesare calculării celor doi parametri a şi b din sistemul de ecuaţii normale pentru o serie de distribuţie de frecvenţe bidimensionale.

Folosirea tabelului de corelaţie pentru calculul produselor de frecvenţă prezintă avantajul că se pot confrunta sumele obţinute pe orizontală cu cele obţinute pe verticală. Astfel, totalul coloanei 9 ($\Sigma y f_y =$ $= 19\,435$) trebuie să fie egal cu totalul rîndului 12. La calculul produselor de frecvenţă trebuie să se ţină seama de toate valorile şi, respectiv, de

Scanned with OKEN Scanner

Tabelul 9.9

Calculul legăturii liniare simple într-un tabel de corelație

(date convenționale)

Nr. crt.	x / y	225	275	325	375	425	475	525	f_y	yf_y	y^2	y^2f_y	xf_y	x^2f_x	xyf_{xy}
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	115	6	1	—	—	—	—	—	7	805	13 225	92 575	1 625	379 375	186 875
2	105	7	9	2	—	—	—	—	18	1 890	11 025	198 450	4 700	1 246 250	493 500
3	95	5	18	5	2	—	—	—	30	2 850	9 025	270 750	8 450	2 423 750	803 250
4	85	1	9	19	17	2	—	—	48	4 080	7 225	346 800	16 100	5 490 000	1 372 430
5	75	2	2	15	25	20	13	2	79	5 925	5 625	444 375	30 975	12 449 375	2 217 125
6	65	—	—	2	6	11	19	15	53	3 445	4 425	223 925	24 475	11 463 125	1 400 875
7	55	—	—	—	—	—	2	6	8	440	3 025	24 200	4 100	2 105 000	237 750
8	f_x	21	39	43	50	33	34	23	243	19 435	—	1 601 075	90 425	35 556 875	6 711 805
9	xf_x	4 725	10 725	13 975	18 750	14 025	16 150	12 075	90 425						
10	x^2	50 625	75 625	105 625	140 625	180 625	225 625	275 625	—						
11	x^2f_x	1 063 125	2 949 375	4 541 875	7 031 250	5 960 625	7 671 250	6 339 375	35 556 875						
12	yf_y	2 135	3 685	3 555	3 900	2 385	2 320	1 455	19 435						
13	y^2f_y	220 125	351 675	297 275	396 850	173 425	159 450	92 775	1 601 075						
14	xyf_y	480 875	1 013 305	1 155 375	1 462 500	1 017 625	788 000	794 125	4 711 805						

Scanned with OKEN Scanner

toate frecvenţele care se întîlnesc pe acelaşi rînd sau pe aceeaşi coloană. Astfel, coloana 9 se obţine înmulţind datele din coloana 0 cu cele din coloana 8, datele din coloana 11, făcînd produsul dintre datele din coloana 10 cu cele din coloana 7 pentru fiecare rînd în parte. Pentru produsele de frecvenţe din coloanele 12, 13, şi 14, se ţine seama de distribuţia lui x pentru fiecare rînd în parte. De exemplu, pentru primul rînd valorile individuale ale lui x sînt: 225 şi 275, iar frecvenţele sînt: 6 şi 1. Pe baza acestei distribuţii se vor obţine:

— pentru coloana 12: $(225 \cdot 6) + (275 \cdot 1) = 1\,625$;

— pentru coloana 13: $(50\,625 \cdot 6) + (75\,625 \cdot 1) = 379\,375$;

— pentru coloana 14: $(225 \cdot 115 \cdot 6) + (275 \cdot 115 \cdot 1) = 186\,875$.

În mod analog se vor stabili produsele de frecvenţe pentru completarea întregii coloane.

La completarea rîndurilor se procedează astfel: pentru rîndurile care se referă la valorile lui x se obţine:

— pentru rîndul 9 produsele de frecvenţe dintre datele de pe rîndul 8 şi valorile lui x din capetele coloanelor, iar pentru rîndul 11 produsele de frecvenţă dintre valorile înscrise pe rîndul 10 şi cele înscrise pe rîndul 8.

Pentru completarea datelor de pe rîndurile 12—14 este necesar să se ţină seama de distribuţia de frecvenţe a valorilor lui y, stabilită pentru fiecare coloană în parte.

Astfel, pentru coloana 1 distribuţia de frecvenţă a lui y va fi:

y	115	105	95	85	75
f_y	6	7	5	1	2

Pe baza acestei distribuţii se vor obţine următoarele produse de frecvenţă pentru rubricile formate din întretăierea coloanei 1 cu rîndurile 12—14:

— rîndul 12: $(115 \cdot 6) + (105 \cdot 7) + (95 \cdot 5) + (85 \cdot 1) + (75 \cdot 2) = 2\,135$;

— rîndul 13: $(13\,225 \cdot 6) + (11\,025 \cdot 7) + (9\,025 \cdot 5) + (7\,225 \cdot 1) + (5\,625 \cdot 2) = 220\,125$;

— rîndul 14: $(225 \cdot 115 \cdot 6) + (225 \cdot 105 \cdot 7) + (225 \cdot 95 \cdot 5) + (225 \cdot 85.1) + (225 \cdot 75 \cdot 2) = 480\,875$.

Avînd calculate toate datele necesare stabilirii sistemului de ecuaţii normale pentru o distribuţie bidimensională de frecvenţă, pe baza tabelului de corelaţie, se obţine:

$$243a + 90\,425\,b = 19\,435;$$

$$90\,425a + 35\,556\,875\,b = 6\,711\,805.$$

Scanned with OKEN Scanner

Rezolvînd sistemul de ecuaţii normale se va proceda ca şi în cazul precedent: se vor determina valorile celor doi parametri a şi b şi, pe baza lor, se va stabili ecuaţia dreptei, după care se va face ajustarea seriei privind distribuţia costului la un ha.a.n. Avînd valorile ecuaţiilor de regresie se va stabili apoi raportul de corelaţie care va măsura gradul de intensitate a legăturii dintre cele două caracteristici corelate.

9.3.1.3. COEFICIENTUL DE CORELAŢIE

În cazul corelaţiei liniare raportul de corelaţie se transformă în coeficient de corelaţie. Determinarea coeficientului de corelaţie liniară simplă se bazează pe distribuţia abaterilor celor două variabile x şi y.

Dacă între cele două variabile există legătură statistică, atunci unităţile purtătoare a celor două variabile vor trebui să aibă o poziţie asemănătoare faţă de valoarea lor medie, deci, vor prezenta abateri perechi de intensităţi asemănătoare şi de acelaşi sens, dacă legătura este directă, şi de sens contrar, dacă legătura este de sens invers.

Aceste abateri se raportează la valoarea abaterii tip a caracteristicii respective, care sintetizează toate abaterile individuale ale caracteristicii:

$$\left(\frac{x_i - \bar{x}}{\sigma_x}\right) \text{ şi } \left(\frac{y_i - \bar{y}}{\sigma_y}\right).$$

Produsul mediu al acestor abateri normate dă tocmai formula de calcul al coeficientului de corelaţie liniară simplă (r_{xy}):

$$r_{xy} = \frac{\Sigma\left(\frac{x_i - \bar{x}}{\sigma_x}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right)}{n} \tag{9.22}$$

este egal cu:

$$r_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x \sigma_y} \tag{9.23}$$

Coeficientul de corelaţie liniară simplă poate să ia valori între -1 şi 0 şi între 0 şi $+1$.

Între -1 şi 0, legătura dintre cele două variabile este de sens invers, iar intensitatea legăturii se apreciază în funcţie de mărimea coeficientului de corelaţie; dacă valoarea sa se apropie de 0, atunci fenomenele sînt independente sau tind către independenţă; dacă se apropie de -1, atunci legătura este foarte strînsă şi de sens invers.

Între 0 şi $+1$, legătura dintre cele două variabile este directă şi este cu atît mai intensă, cu cît se apropie de 1, care indică legătura funcţională.

Semnul coeficientului de corelaţie este determinat de datele din numărător, după modul de asociere a abaterilor perechi. Dacă predomină abaterile perechi de acelaşi semn, legătura va fi directă; dacă predomină abaterile de sens contrar, atunci legătura este de sens invers.

Scanned with OKEN Scanner

Folosind acelaşi exemplu luat pentru calcularea raportului de corelaţie (vezi tabelul 9.3) se va determina şi coeficientul de corelaţie. Pentru aceasta se va folosi un tabel de calcul în care se vor trece elemente necesare calculării coeficientului de corelaţie :

Tabelul 9.10

Calculul coeficientului de corelaţie

x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
1	2	3	4	5	6	7
142	123	−29,7	882,09	−16,7	278,89	495,99
149	126	−22,7	515,29	−13,7	187,69	310,99
157	128	−14,7	216,09	−11,7	136,89	171,99
169	134	− 2,7	7,29	− 5,7	32,49	15,39
181	143	− 9,3	86,49	3,3	10,89	30,69
194	210	22,3	457,29	17,3	299,29	385,79
210	167	38,3	1 466,89	27,3	745,29	1 045,59
1 202	978	—	3 631,43	—	1 691,43	2 456,43
Σx	Σy		$\Sigma(x-\bar{x})^2$		$\Sigma(y-\bar{y})^2$	$\Sigma(x-\bar{x})(y-\bar{y})$

Pentru a calcula coeficientul de corelaţie trebuie să se calculeze, în primul rînd, cele două medii şi cele două abateri medii pătratice pentru cele două variabile x şi y :

$$\bar{x} = \frac{\Sigma x}{n} = \frac{1\,202}{7} = 171{,}7\,; \quad \bar{y} = \frac{\Sigma y}{n} = \frac{978}{7} = 139{,}7\,;$$

$$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{3\,631{,}43}{7}} = 22{,}8\,;$$

$$\sigma_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}} = \sqrt{\frac{1\,691{,}43}{7}} = 15{,}6.$$

Datele obţinute se vor înlocui în formula coeficientului de corelaţie :

$$r_{x,y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\sigma_x\,\sigma_y} = \frac{2\,456{,}43}{7 \cdot 22{,}8 \cdot 15{,}6} = 0{,}986$$

După cum s-a arătat, în cazul corelaţiei liniare simple raportul de corelaţie se transformă în coeficient de corelaţie, astfel încît valorile celor doi indicatori sînt egale. În exemplul luat, raportul de corelaţie a fost egal cu 0,988, iar coeficientul de corelaţie cu 0,986, diferenţa dintre cele două valori este nesemnificativă, ea provenind din eventualele rotunjiri de calcule.

În teoria şi practica statistică, la calculul coeficientului de corelaţie se folosesc şi unele formule de calcul simplificat, utilizînd datele deja calculate pentru determinarea ecuaţiilor de regresie. Cel mai frecvent se

Scanned with OKEN Scanner

folosește formula de calcul simplificat obținută prin înlocuirea mediilor și abaterilor medii pătratice cu formulele lor de calcul. Prin simplificare se obține următoarea formulă de calcul simplificat al coeficientului de corelație simplă:

$$r_{xy} = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}. \tag{9.24}$$

Examinînd elementele din formula de calcul simplificat al coeficientului de corelație liniară simplă se observă că, cu excepția Σy^2, ele se găsesc în formulele de calcul ale celor doi parametri din ecuația funcției liniare. Acesta este motivul pentru care în tabelul 9.4 s-a trecut în plus și o coloană pentru y^2.

Folosind datele din tabelul 9.4 se obține:

$$r_{xy} = \frac{7 \cdot 170\,393 - 1\,202 \cdot 978}{\sqrt{(7 \cdot 210\,072 - 1\,202)^2 (7 \cdot 138\,322 - 978^2)}} = 0{,}985.$$

În cazul distribuțiilor de frecvență formula coeficientului de corelație liniară simplă va fi:

$$r_{xy} = \frac{\Sigma(x - \bar{x})(y - \bar{y}) f}{\Sigma f \sigma_x \sigma_y}. \tag{9.25}$$

Iar pentru calculul simplificat:

$$r_{x,y} = \frac{\Sigma f(\Sigma xy f_{xy}) - (\Sigma x f_x)(\Sigma y f_y)}{\sqrt{[\Sigma f \Sigma x^2 f_x - (\Sigma x f_x)^2][\Sigma f \Sigma y^2 f_y - (\Sigma y f_y)^2]}} \tag{9.26}$$

9.3.1.4. RELAȚII ÎNTRE COEFICIENTUL DE CORELAȚIE ȘI PARAMETRII ECUAȚIEI DE REGRESIE LINIARĂ SIMPLĂ

În cazul legăturii liniare între coeficientul de corelație liniară simplă, coeficientul de regresie și cele două abateri medii pătratice ale lui x și y există o relație bine determinată și anume:

$$b = r_{xy} \frac{\sigma_y}{\sigma_x}. \tag{9.27}$$

Această relație se poate folosi atunci cînd există certitudinea că legătura este de forma liniei drepte și pentru a evita calculul direct al parametrilor funcției de regresie se poate trece direct la ajustare. Pentru aceasta se folosește sistemul de ecuații normale în care valorile variabilelor x

Scanned with OKEN Scanner

și y se înlocuiesc cu abaterile lor față de media respectivă. Sistemul de ecuații normale va fi :

$$na + b\Sigma(x - \bar{x}) = \Sigma (y - \bar{y})$$

$$a\Sigma (x - \bar{x}) + b\Sigma (x - \bar{x})^2 = \Sigma (x - \bar{x}) (y - \bar{y}),$$

în care :

$$\Sigma (x - \bar{x}) = 0 ;$$

$$\Sigma (y - \bar{y}) = 0.$$

Deci, primul sistem se simplifică astfel :

$$na = 0$$

$$\Sigma (x - \bar{x})^2 = \Sigma (x - \bar{x})(y - \bar{y}),$$

de unde :

$$a = 0 ;$$

$$b = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2}. \qquad (9.28)$$

Dacă se ia în considerație că pe grafic parametrul a este egal cu punctul ale cărui coordonate sînt cele două valori medii $(\bar{x}, \bar{y})$, atunci ecuația dreptei care trece prin acest punct este :

$$\bar{Y}_x = a + b\bar{x}.$$

În această ecuație $\bar{Y}_x$ se poate înlocui cu $\bar{y}$, deoarece $\Sigma y = \Sigma Y_x$ deci $\bar{y} = \bar{Y}_x$ și se obține :

$$\bar{y} = a + b\bar{x}.$$

Cele două medii $(\bar{x}, \bar{y})$ fiind cunoscute de la calculul abaterilor medii pătratice, înseamnă că se va putea calcula și valoarea lui a din ecuația funcției de regresie rectilinie :

$$a = \bar{y} - b\bar{x}.$$

Determinînd cei doi parametri se poate face ajustarea în funcție de valorile individuale ale caracteristicii x și calcula apoi raportul de corelație pentru care nu este necesar să se calculeze decît dispersia față de linia de regresie, dispersia totală fiind cunoscută de la calculul coeficientului de corelație liniară simplă.

Dacă legătura dintre cele două variabile este liniară, atunci coeficientul de corelație liniară simplă este o medie geometrică a coeficienților de

Scanned with OKEN Scanner

regresie corespunzători dependenţei reciproce a celor două variabile x şi y :

$$r_{xy} = \sqrt{b_{y/x} \cdot b_{x/y}}. \tag{9.29}$$

Pentru dependenţa reciprocă, folosind metoda celor mai mici pătrate există cele două sisteme de ecuaţii, în care pe rînd cele două variabile sînt considerate în funcţie de cealaltă :

— primul sistem pentru y dependent de x :

$$na + b\Sigma x = \Sigma y$$

$$a\Sigma x + b\Sigma x^2 = \Sigma xy$$

— al doilea sistem pentru x dependent de y :

$$na + b\Sigma y = \Sigma x \tag{9.30}$$

$$a\Sigma y + b\Sigma y^2 = \Sigma xy.$$

Calcularea coeficientului de regresie $b\, y/x$ se face folosind relaţia (9.10), iar a coeficientului $b\ x/y$ se obţine astfel :

$$b_{x/y} = \frac{\begin{vmatrix} n & \Sigma x \\ \Sigma y & \Sigma xy \end{vmatrix}}{\begin{vmatrix} n & \Sigma y \\ \Sigma y & \Sigma y^2 \end{vmatrix}} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma y^2 - (\Sigma y)^2}. \tag{9.31}$$

Înlocuind în relaţia (9.29) cei doi coeficienţi de regresie (9.10) şi (9.31) cu formulele de calcul se obţine :

$$r_{xy} = \sqrt{\frac{[n\Sigma xy - \Sigma x \Sigma y][n\Sigma xy - (\Sigma x)(\Sigma y)]}{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}.$$

Avînd la numărător aceeaşi expresie la pătrat, se scoate de sub radical şi se obţine :

$$r_{x,y} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}.$$

Se observă că s-a ajuns la aceeaşi formulă ca şi a coeficientului de corelaţie liniară folosind calculul simplificat (6.24).

După cum s-a mai arătat, în cazul legăturii liniare coeficientul de corelaţie simplă este egal cu raportul de corelaţie simplă. În acest sens, coeficientul de corelaţie liniară simplă poate fi considerat ca un test de verificare a liniarităţii legăturii. Dacă nu există certitudinea că legătura este de

Scanned with OKEN Scanner

formă rectilinie, este necesar să se calculeze şi raportul de corelaţie a cărui formulă depinde de forma de legătură. Dacă se obţine pentru raportul de corelaţie o valoare diferită faţă de coeficientul de corelaţie liniară simplă, înseamnă că legătura este curbilinie şi trebuie să se reţină numai valoarea raportului de corelaţie care îndeplineşte cel mai bine condiţia de minim.

9.3.2. Corelaţia simplă curbilinie

În practica statistică se întîlnesc uneori şi legături care nu se realizează de formă liniară. În acest caz se pot întîlni forme variate de legături curbilinii specifice relaţiilor de dependenţă dintre caracteristica factorială aleasă şi cea rezultativă.

Aceste legături neliniare se întîlnesc mai frecvent sub forma unei parabole de gradul 2, a hiperbolei şi a funcţiei exponenţiale. În acest caz este necesar ca ecuaţia funcţiei de estimare a caracteristicii rezultative să corespundă formei obiective de manifestare a legăturii.

În cazul legăturii neliniare este necesar ca, pe baza graficului, să se aleagă acea funcţie care, trasată vizual, să prezinte abateri minime de la linia valorilor empirice înscrise în cîmpul de corelaţie. Cînd nu se poate desprinde cu claritate forma de realizare a legăturii dintre x şi y este necesar să se calculeze mai multe ecuaţii de estimare, cu care să se ajusteze datele şi să se aleagă apoi aceea care dă o dispersie minimă faţă de valorile teoretice ale caracteristicii rezultative în funcţie de factorul ales.

9.3.2.1. DETERMINAREA ECUAŢIILOR DE REGRESIE

Dacă pe grafic se apreciază că legătura statistică este de forma unei parabole de gradul 2, atunci ecuaţia medie de estimare este:

$$\bar{Y}_x = a + bx + cx^2. \tag{9.32}$$

De remarcat este faptul că, în practica statistică, se pot întîlni legături de forma unei parabole de gradul 2, care să prezinte puncte de minim sau maxim, sau numai secţiuni de parabole crescătoare sau descrescătoare. În mod corespunzător se vor obţine valori pozitive şi valori negative sau numai valori într-un singur sens pentru parametrii a, b, c. Pentru cazul cînd apare un punct de maxim sau de minim este necesar să se aprecieze, pe baza funcţiei de estimare, care este punctul de la care legătura începe să fie nesemnificativă.

Punînd aceeaşi condiţie, ca suma pătratelor abaterilor termenilor seriei de la valorile teoretice să fie minimă, se obţine:

$$\Sigma(y - Y_x)^2 = \text{minim},$$

în care:

$$Y_x = (a + bx + cx^2),$$

Scanned with OKEN Scanner

de unde:

$$\Sigma[(y - (a + bx + cx^2)]^2 = \min.$$

Aplicînd procedeul multiplicării cu coeficienţii celor trei parametri a, b, c se va obţine un sistem format dintr-un număr egal cu numărul parametrilor funcţiei:

$$a + bx + cx^2 = y \ (1)$$

$$a + bx + cx^2 = y \ (x)$$

$$a + bx + cx^2 = y \ (x^2)$$

Multiplicînd ecuaţiile, însumînd pentru cele n observaţii:

$$na + \Sigma bx + \Sigma cx^2 = \Sigma y$$

$$\Sigma ax + \Sigma bx^2 + \Sigma cx^3 = \Sigma xy$$

$$\Sigma ax^2 + \Sigma bx^3 + \Sigma cx^4 = \Sigma x^2 y$$

Scoţînd necunoscutele (a, b, c) în factor se obţine:

$$na + b\Sigma x + c\Sigma x^2 = \Sigma y$$

$$a\Sigma x + b\Sigma x^2 + c\Sigma x^3 = \Sigma xy \qquad (9.33)$$

$$a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 = \Sigma x^2 y$$

Rezolvînd sistemul de ecuaţii normale prin metoda determinanţilor se calculează valoarea celor trei parametri şi, în funcţie de valoarea individuală a lui x, se ajustează valorile caracteristicii rezultative.

În acest caz este nevoie de tabel de calcul (tabelul 9.11).

Tabelul 9.11

Modelul tabelului de calcul folosind ecuaţia de estimare de forma unei parabole de gradul 2

Nr. crt.	x	y	x^2	x^3	x^4	xy	x^2y	$Y = a + bx + cx^2$
1	x_1	y_1	x_1^2	x_1^3	x_1^4	x_1y_1	$x_1^2\,y$	$Y_{x_1} = a + bx_1 + cx_n^2$
2	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮								
n	x_n	y_n	x_n^2	x_n^3	x_n^4	x_ny_n	$x_n^2\,y$	$Y_{x_n} = a + bx_n + cx_1^2$
Total	Σx	Σy	Σx^2	Σx^3	Σx^4	Σxy	Σx_y^2	$\Sigma Y_x = \Sigma y$

Scanned with OKEN Scanner

Cînd pe grafic se constată o corelaţie de forma unei hiperbole trebuie să se aplice metoda celor mai mici pătrate în care ecuaţia de estimare este de forma :

$$Y_x = a + \frac{1}{x} b \tag{9.34}$$

Punînd aceeaşi condiţie :

$$\Sigma(y - Y_x)^2 = \min$$

$$\Sigma \left[\left(y - \left(a + b \frac{1}{x} \right) \right] \right]^2 = \min$$

După acelaşi procedeu sistemul de ecuaţii normale va fi format din două ecuaţii, care se obţin prin aplicarea coeficienţilor celor doi parametri ai ecuaţiei funcţiei :

$$a + \frac{1}{x} b = y \ (1)$$

$$a + \frac{1}{x} b = y \left(\frac{1}{x} \right)$$

Multiplicînd şi însumînd se obţine :

$$\Sigma a + \Sigma \frac{1}{x} b = \Sigma y$$

$$\Sigma ax + \Sigma \frac{1}{x^2} b = \Sigma \frac{1}{x} y$$

Scoţînd în factori necunoscutele :

$$na + b \Sigma \frac{1}{x} = \Sigma y$$

$$a\Sigma \frac{1}{x} + b\Sigma \frac{1}{x^2} = \Sigma \frac{1}{x} y \tag{9.35}$$

Rezolvînd sistemul se obţine ecuaţia funcţiei de estimare a caracteristicii rezultative de forma unei hiperbole şi calculînd fiecare ecuaţie de regresie se trece la ajustarea seriei.

Scanned with OKEN Scanner

Tabelul 9.12

Modelul tabelului de calcul folosind ca ecuaţie de estimare hiperbola

Nr. crt.	x	y	$\frac{1}{x}$	$\frac{1}{x^2}$	$\frac{1}{x}y$	$Y_x = a + b\frac{1}{x}$
1 . . . . . n						
Total	Σx	Σy	$\Sigma \frac{1}{x}$	$\Sigma \frac{1}{x^2}$	$\Sigma \frac{1}{x} y$	$\Sigma Y_x = \Sigma y$

Apar şi cazuri cînd legătura dintre cele două variabile este de formă exponenţială.

$$\bar{Y}_x = a \cdot b^x. \tag{9.36}$$

Aplicînd metoda celor mai mici pătrate se obţine:

$$\Sigma(y - Y_x)^2 = \min.$$

$$\Sigma(y - ab^x)^2 = \min.$$

Pentru care sistemul de ecuaţii normale se determină folosind logaritmii celor doi parametri a şi b:

$$\log \bar{Y}_x = \log\ a + x\ \log\ b. \tag{9.37}$$

Deci, s-a ajuns la o ecuaţie de estimare de forma unei linii drepte calculată nu pe baza valorilor empirice ci a logaritmilor lor.

Sistemul de ecuaţii normale va fi cel de la corelaţie liniară simplă, în care valorile celor doi parametri vor fi înlocuite cu valorile logaritmilor lor, iar valorile caracteristicii rezultative y cu valorile logaritmilor acestora:

$$\begin{aligned} n\ \log\ a + \log\ b\Sigma x &= \Sigma\ \log\ y \\ \log\ a\Sigma x + \log\ b\Sigma x^2 &= \Sigma(x\ \log\ y) \end{aligned} \tag{9.38}$$

Scanned with OKEN Scanner

Operaţia de ajustare în acest caz se va face după ce se vor calcula logaritmii ecuaţiilor. Pentru ajustarea după o funcţie exponenţială este necesar următorul tabel de calcul:

Tabelul 9.13

Model de tabel de calcul folosind ca funcţie de estimare ecuaţia funcţiei exponenţiale

Nr. crt.	x	y	$\log y$	x^2	$x \log y$	$\log Y_x = \log a + x \log b$	$Y_x = ab^x$
1	x_1	y_1	$\log y_1$	x_1^2	$x_1 \log y_1$	$\log Y_{x_1} = \log a + x_1 \log b$	$Y_{x_1} = ab^{x_1}$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
n	x_n	y_n	$\log y_n$	x_n^2	$x_n \log y_n$	$\log Y_{x_n} = \log a + x_n \log b$	$Y_{x_n} = ab^{x_n}$
Total	Σx	Σy	$\Sigma \log y$	Σx^2	$\Sigma x \log y$	$\Sigma \log Y_x = \Sigma \log y$	$\Sigma Y_x = \Sigma y$

9.3.2.2. RAPORTUL DE CORELAŢIE

Pentru corelaţia neliniară măsurarea gradului de intensitate a legăturii se face la corelaţia curbilinie *numai* prin raportul de corelaţie:

— pentru legătura sub forma unei parabole de gradul 2:

$$R_{x,y} = \sqrt{1 - \frac{\Sigma(y_i - Y_x)^2}{\Sigma(y_i - \bar{y})^2}} = \sqrt{1 - \frac{\Sigma[y_i - (a + bx + cx^2)]^2}{\Sigma(y_i - \bar{y})^2}}; \quad (9.39)$$

— pentru legătura sub formă de hiperbolă:

$$R_{x,y} = \sqrt{1 - \frac{\Sigma(y_i - Y_x)^2}{\Sigma(y_i - \bar{y})^2}} = \sqrt{1 - \frac{\Sigma\left[\left(y_i - \left(a + \frac{1}{x} b\right)\right]\right]^2}{\Sigma(y_i - \bar{y})^2}}; \quad (9.40)$$

— pentru legătura sub forma unei funcţii exponenţiale:

$$R_{x,y} = \sqrt{1 - \frac{\Sigma(y_i - Y_x)^2}{\Sigma(y_i - \bar{y})^2}} = \sqrt{1 - \frac{\Sigma(y_i - ab^x)^2}{\Sigma(y_i - \bar{y})^2}} \quad (9.41)$$

Pentru calculul raportului de corelaţie este necesară determinarea celor două dispersii stabilite faţă de linia de regresie şi faţă de valoarea medie. Datele necesare calculării lor se înscriu în următorul tabel în care Y_x poate fi:

$$Y_x = a + bx + cx^2; \quad Y_x = a + b\frac{1}{x}; \quad Y_x = ab^x.$$

Scanned with OKEN Scanner

Tabelul 9.14

Modelul tabelului de calcul al raportului de corelaţie

Nr. crt.	y	Y_x	$y - Y_x$	$(y - Y_x)^2$	$(y - \bar{y})$	$(y - \bar{y})^2$
1	y_1	Y_{x_1}	$(y_1 - Y_{x_1})$	$(y_1 - Y_{x_1})^2$	$(y_1 - \bar{y})$	$(y_1 - \bar{y})^2$
2	y_2	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
n	y_n	Y_{x_n}	$(y_n - Y_{x_n})$	$(y_n - Y_{x_n})^2$	$(y_n - \bar{y})$	$(y_n - \bar{y})^2$
Total		—	—	$\Sigma(y_i - Y_x)^2$	—	$\Sigma(y_i - \bar{y})^2$

Măsurarea gradului de interdependenţă dintre cele două variabile supuse corelaţiei se face însă numai după ce s-a verificat, grafic sau prin analiza dispersională, obiectivitatea funcţiei de ajustare care a fost aleasă.

Dacă s-a stabilit că legătura este neliniară apare clar că nu mai are sens calculul coeficientului de corelaţie liniară simplă. Pentru verificarea obiectivităţii funcţiei alească pentru ajustare se folosesc, în acest caz, în special procedeele de analiză dispersională. În cazul folosirii analizei dispersionale verificarea semnificaţiei liniei (curbei) de regresie se face prin descompunerea dispersiei totale în funcţie de forma funcţiei de regresie folosită şi calcularea mărimii dispersiei din interiorul grupelor ca mijloc de estimare a influenţei factorilor reziduali.

Gradele de libertate folosite pentru calculul dispersiei factorului se vor stabili, în acest caz, pornind de la numărul parametrilor funcţiei de estimare şi numărul total al observaţiilor[1].

9.3.3. Corelaţia multiplă în cazul distribuţiilor multidimensionale de formă liniară

Fenomenele sociale de masă — după cum s-a mai arătat — sînt fenomene complexe, care depind de un număr mare de cauze şi condiţii. Forma de manifestare a acestor fenomene este în funcţie de modul de asociere

---

[1] Vezi capitolul „Elemente de analiză dispersională".

$$F_{calc.} = \frac{S^2_{y/x}}{S^2_{y/z}}$$

$S^2_{y/z}$ se calculează cu formula (8.9), iar:

$$S^2_{y/x} = \frac{\Sigma(y - Y_x)^2}{r - k},$$

în care:

r reprezintă numărul grupelor iar K numărul parametrilor funcţiei de ajustare.

Dacă $F_{calc.} < F_{tab.}$ atunci se verifică ipoteza funcţiei de ajustare folosită.

Scanned with OKEN Scanner

a cauzelor care le determină, ceea ce înseamnă că mai multe cauze pot duce la producerea unui anumit efect. Cel mai frecvent, însă, în modul de variaţie al caracteristicii rezultative se manifestă nu numai influenţa fiecăruia dintre factorii determinanţi, ci mai ales interacţiunea dintre ei. Aceasta face ca gradul de influenţă a diferiţilor factori să fie variabil de la o unitate la alta, iar pe ansamblu influenţa unui factor să se regăsească prin intermediul unui alt factor cu care se găseşte în interdependenţă.

9.3.3.1. DETERMINAREA ECUAŢIILOR DE REGRESIE

În cazul fenomenelor de masă se consideră variabila dependentă sau rezultativă ca fiind o funcţie de mai multe variabile : $y = f(x_1, x_2 \ldots, x_n)$, în care variabilele x sînt caracteristicile factoriale care determină într-o măsură mai mare sau mai mică variaţia caracteristicii rezultative y. Dacă legătura dintre fiecare factor şi variabila rezultativă este de formă liniară, atunci ecuaţia de estimare va fi :

$$Y_{x_1, x_2, \ldots, x_n} = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n, \qquad (9.42)$$

în care :

a_0 reprezintă parametrul care exprimă factorii neînregistraţi, consideraţi cu acţiune constantă, în afara celor consideraţi drept caracteristici factoriale ;

$a_1, a_2, \ldots, a_n$ — coeficienţii de regresie care arată cu cît se modifică caracteristica rezultativă dacă caracteristica factorială respectivă se modifică cu o unitate ;

$x_1, x_2, \ldots, x_n$ — caracteristicile factoriale incluse în raportul de interdependenţă.

Calcularea parametrilor $a_0, a_1, a_2, \ldots, a_n$ se face prin aplicarea metodei celor mai mici pătrate, punînd condiţia ca suma abaterilor termenilor empirici de la linia de regresie, ridicate la pătrat, să fie minimă :

$$\Sigma(y_i - Y_{x_1, x_2, \ldots, x_n})^2 = \min.$$

adică :

$$\Sigma[(y_i - (a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n)]^2 = \min.$$

Pentru a afla valoarea acestor parametri este necesar să se stabilească sistemul corespunzător de ecuaţii normale.

Pentru aceasta se va înmulţi, pe rînd, ecuaţia funcţiei de regresie multiplă cu coeficienţii parametrilor $a_{0,1,2,\ldots,n}$ şi se va însuma pentru toate valorile individuale ale caracteristicilor supuse analizei de corelaţie :

$$\begin{aligned}
a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n &= y \ (1) \\
a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n &= y \ (x_1) \\
a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n &= y \ (x_2) \\
\vdots \quad \vdots \quad \vdots \quad & \quad \vdots \quad \vdots \\
a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n &= y \ (x_n)
\end{aligned}$$

Scanned with OKEN Scanner

Efectuînd înmulţirile şi însumînd ecuaţiile pentru toate valorile individuale ale caracteristicilor şi scoţînd parametrii funcţiei în factori, se obine sistemul de ecuaţii normale :

$$\begin{aligned}
&na_0 + a_1\Sigma x_1 + a_2\Sigma x_2 + \ldots + a_n\Sigma x_n = \Sigma y \\
&a_0\Sigma x_1 + a_1\Sigma x_1^2 + a_2\Sigma x_1x_2 + \ldots + a_n\Sigma x_1x_n = \Sigma x_1y \qquad (9.43)\\
&a_0\Sigma x_2 + a_1\Sigma x_1x_2 + a_2\Sigma x_2^2 + \ldots + a_n\Sigma x_2x_n = \Sigma xy \\
&\vdots \\
&a_0\Sigma x_n + a_1\Sigma x_1x_2 + a_2\Sigma x_2x_n + \ldots + a_n\Sigma a_nx_n = \Sigma x_ny
\end{aligned}$$

Spre exemplificare se va utiliza o corelaţie multiplă în care se vor lua numai două caracteristici factoriale şi una rezultativă. Pentru aceasta se va calcula corelaţia dintre dinamica înzestrării tehnice a muncii, productivitatea muncii şi a produsului global al industriei în perioada 1970—1976.

Tabelul 9.15

Dinamica înzestrării tehnice a muncii, a productivităţii muncii şi a produsului global din industrie, în perioada 1970—1976

(1965 = 100)

Anii	Înzestrarea tehnică a muncii * (x_1)	Productivitatea muncii (x_2)	Produsul global (y)
A	1	2	3
1970	151	142	172
1971	161	149	191
1972	168	157	212
1973	176	169	242
1974	187	181	274
1975	207	194	303
1976	221	210	336

* Calculat ca raport între dinamica fondurilor fixe şi dinamica personalului încadrat în industrie

Sursa: Anuarul Statistic al R.S. România, D.C.S., 1977, p. 52, 56, 68, 69, 137.

După cum se ştie, dinamica produsului global al industriei depinde atît de dinamica înzestrării tehnice, cît şi de dinamica productivităţii muncii. Astfel, pe măsură ce creşte gradul de înzestrare tehnică a muncii, trebuie să crească productivitatea muncii, ceea ce asigură în continuare creşterea producţiei. Pentru realizarea relaţiei de interdependenţă dintre cele trei variabile statistice, înzestrarea tehnică a muncii poate fi considerată ca o primă caracteristică factorială (x_1) ce influenţează asupra dinamicii produsului global (y), iar productivitatea muncii ca cea de-a doua caracteristică factorială (x_2).

Avînd două caracteristici factoriale ecuaţia funcţiei de regresie liniară multiplă va fi :

$$\bar{Y}_{x_1, x_2} = a_0 + a_1x_1 + a_2x_2 \qquad (9.44)$$

al cărei sistem de ecuaţii normale pentru aflarea celor trei parametri se prezintă astfel:

$$\begin{aligned} n\Sigma a_0 + a_1\Sigma x_1 + a_2\ \Sigma\ x_2 &= \Sigma y\,; \\ a_0\Sigma x_1 + a_1\Sigma x_1^2 + a_2\Sigma x_1 x_2 &= \Sigma x_1 y\,; \\ a_0\Sigma x_2 + a_1\Sigma x_1 x_2 + a_2\Sigma x_2^2 &= \Sigma x_2 y. \end{aligned} \qquad (9.45)$$

Folosind datele din tabelul 9.16, se obţine sistemul de ecuaţii necesar calculării celor trei parametri ai ecuaţiei de regresie multiplă:

$$7\ a_0 + 1\,217\ a_1 + 1\,202\ a_2 = 1\,730\,;$$

$$1\,271\ a_0 + 234\,581\ a_1 + 222\,966\ a_2 = 323\,146\,;$$

$$1\,730\ a_0 + 222\,966\ a_1 + 210\,072\ a_2 = 306\,001.$$

Rezolvînd sistemul de ecuaţii se obţine:

$$a_0 = -161{,}6\,;$$

$$a_1 = 2{,}21\,;$$

$$a_2 = 0{,}044.$$

Deci, ecuaţia medie de tendinţă va fi:

$$\bar{Y}_{x_1, x_2} = -161{,}6 + 2{,}21\ x_1 + 0{,}644\ x_2.$$

Interpretînd statistic cei doi coeficienţi de regresie liniară multiplă se poate afirma că perioadei celor şapte ani (1970—1976) la o creştere medie anuală de 1% a înzestrării tehnice a muncii în interdependenţă cu productivitatea muncii îi corespunde o creştere a produsului global cu 0,27% pe seama creşterii înzestrării tehnice şi cu 2,35% pe seama creşterii productivităţii muncii.

Influenţa mai mare a celui de-al doilea factor asupra caracteristicii rezultative este justificată prin faptul că înzestrarea tehnică a muncii determină în mod hotărîtor nivelul productivităţii muncii.

Cunoscînd cei trei parametri ai funcţiei de ajustare se calculează pentru fiecare an ecuaţia de regresie (vezi tabelul 9.16, col. 10), iar pentru verificarea calculului se aplică relaţia $\Sigma_y = \Sigma Y_{x_1, x_2}$. (Diferenţa nesemnificativă dintre suma valorilor empirice şi a celor ajustate este rezultatul rotunjirii valorii celor trei parametri).

9.3.3.2. RAPORTUL DE CORELAŢIE

Pentru măsurarea gradului de intensitate a corelaţiei se foloseşte şi aici raportul de corelaţie.

Raportul de corelaţie multiplă se va calcula ca şi în cazul corelaţiei simple în funcţie de greutatea specifică a dispersiei provocată de factorii înregistraţi ($\sigma^2_{y/x_1,x_2}$) faţă de dispersia totală a caracteristicii rezultative (σ^2_y) sau folosind relaţia dintre cele trei dispersii ($\sigma^2_y = \sigma^2_{y/x_1x_2} + \sigma_{y/r}$).

$$R_{x_1,x_2/y} = \sqrt{1 - \frac{\Sigma(y_i - Y_{x_1x_2})^2}{\Sigma(y_i - \bar{y})^2}}\,. \qquad (9.46)$$

Scanned with OKEN Scanner

Tabelul 9.16

Ajustarea pe baza ecuației liniare multiple (x_1, x_2).

x_1	x_2	y	x_1^2	x_2^2	x_1x_2	x_1y	x_2y	y^2	Y_{x_1, x_2}
1	2	3	4	5	6	7	8	9	10
151	142	172	22 801	20 164	21 442	25 972	24 424	29 584	178
161	149	191	25 921	22 201	23 989	30 751	28 459	36 481	201
168	157	212	28 224	24 649	26 376	35 616	33 284	44 944	217
176	169	242	30 976	28 561	29 744	42 592	40 898	58 564	235
187	181	274	34 969	32 761	33 847	51 238	49 594	75 076	260
207	194	303	42 849	37 636	40 158	62 721	58 782	91 809	305
221	210	336	48 841	44 100	46 410	74 256	70 560	112 896	330
1 271	1 202	1 730	234 581	210 072	221 966	323 146	306 001	449 354	1 726
Σx_1	Σx_2	Σy	Σx_1^2	Σx_2^2	Σx_1x_2	Σx_1y	Σx_2y	Σy^2	$\Sigma Y_m = \Sigma y$

$$\tilde{Y}_{x_1x_2} = -161{,}6 + 2{,}22x_1 + 0{,}044x_2.$$

Scanned with OKEN Scanner

Raportul de corelaţie multiplă poate lua valori între —1 şi 0, între 0 şi +1. Raportul de corelaţie multiplă are valoarea cea mai ridicată în raport cu indicii de corelaţie simplă, deoarece el întruneşte atît influenţa fiecărui factor, cît şi influenţa interacţiunii dintre ei. Deci, cu cît se iau în considerare mai mulţi factori, cu atît valoarea raportului de corelaţie multiplă trebuie să fie mai mare. Teoretic se admite că dacă toţi factorii s-ar putea exprima numeric, atunci raportul de corelaţie multiplă ar trebui să aibă valoarea 1, arătînd dependenţa funcţională dintre toţi factorii săi determinanţi luaţi la un loc şi nivelul său, iar coeficienţii de regresie ar arăta cu ce valoare în plus sau în minus s-ar modifica caracteristica rezultativă ca urmare a modificării cu o unitate a factorului respectiv. Deci, în acest caz ecuaţia de regresie ar fi egală cu valoarea empirică a caracteristicii factoriale calculată în funcţie de mărimea tuturor factorilor săi determinanţi, iar termenul liber ar fi egal cu zero:

$$Y_{x_1, x_2, \ldots, x_n} = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n.$$

În realitate, însă, neputînd identifica în întregime toţi factorii, nefiind posibilă întotdeauna exprimarea lor numeric, valoarea liniei de regresie multiplă se va abate mai mult sau mai puţin de la valorile reale ale termenilor seriei, ca urmare a influenţei acelor factori neînregistraţi şi incluşi în valoarea termenului liber a_0.

Pentru a calcula raportul de corelaţie multiplă este necesar să se determine cele două dispersii: dispersia valorilor caracteristicii y de la linia lor de regresie multiplă şi dispersia aceleiaşi valori de la media lor (tabelul 9.17).

Tabelul 9.17

Calculul dispersiei valorilor empirice faţă de valorile ecuaţiilor de regresie şi faţă de valoarea medie

Anii	Indicii produsului global al industriei 1965 = 100 y	Valorile ecuaţiilor de regresie $Y_{x_1 x_2}$	Abaterile valorilor y de la media lor $y - \bar{y}$	Pătratul abaterilor $(y - \bar{y})^2$	Abaterile valorilor y de la ecuaţia de regresie $y - Y_{x_1, x_2}$	Pătratul abaterilor faţă de ecuaţia de regresie $(y - Y_{x_1, x_2})^2$
A	1	2	3	4	5	6
1970	172	178	—75	5 625	— 6	36
1971	191	201	—56	3 136	—10	100
1972	212	217	—35	1 225	— 5	25
1973	242	235	— 5	25	7	49
1974	274	260	27	729	14	196
1975	303	305	56	3 136	2	4
1976	336	330	89	7 921	6	36
—	1 730	—	—	21 797	—	446

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{1\,730}{7} = 247{,}1;$$

$$R_{x_1, x_2, y} = \sqrt{1 - \frac{\Sigma (y_i - Y_{x_1, x_2})^2}{\Sigma (y_i - \bar{y})^2}} = \sqrt{1 - \frac{446}{21\,797}} = 0{,}99.$$

Scanned with OKEN Scanner

Deci, între cele trei caracteristici, în perioada 1970—1976 a existat o legătură foarte strînsă, dinamica produsului global al industriei depinzînd într-un grad foarte mare de cei doi factori : gradul de înzestrare tehnică a muncii şi productivitatea muncii din întreprinderile industriale. Semnul raportului de corelaţie este dat de coeficienţii de regresie. Pentru exemplul utilizat cei doi coeficienţi de corelaţie sînt pozitivi, deci, legătura dintre cele trei caracteristici a fost de tip liniar şi de sens direct. În cazul în care coeficienţii de regresie au semne diferite, atunci se consideră legăturile simple cu fiecare factor în parte, pentru interpretarea sensului legăturii.

În cazul analizat, fiind vorba de un indicator cu un grad mare de sintetizare a diferiţilor factori — produsul global industrial — deşi coeficientul de corelaţie are o valoare de necontestat în procesul de analiză a gradului de legătură dintre cele trei caracteristici, el singur nu este suficient pentru întreaga muncă de analiză. În cazul analizei de corelaţie a producţiei, deosebit de importantă apare însăşi interpretarea ecuaţiilor de regresie, cu alte cuvinte, valorile ajustate ale termenilor y, calculate în funcţie de factorii respectivi şi compararea lor cu nivelul real atins în fiecare an în parte. Cu cît valoarea raportului este mai mare, cu atît influenţa factorilor neînregistraţi este mai puternică la nivelul unităţilor la care se referă ecuaţia de estimare.

Aceasta demonstrează că şi în interpretarea corelaţiei dintre fenomene trebuie folosiţi toţi indicatorii care permit analiza modului cum se realizează legătura dintre ele.

În cazul corelaţiei liniare, raportul de corelaţie multiplă se transformă în coeficient de corelaţie multiplă.

Coeficientul de corelaţie multiplă poate fi dedus din coeficienţii de corelaţie simplă. După cum se ştie, în cazul corelaţiei liniare simple, raportul de corelaţie este egal cu coeficientul de corelaţie ($R_{x,y} = r_{x,y}$).

În cazul corelaţiei multiple, raportul de corelaţie liniară sintetizează toate legăturile liniare simple. Dacă factorii sînt independenţi între ei, atunci raportul de determinaţie multiplă este egal cu suma rapoartelor de determinaţie simplă. De exemplu, pentru doi factori există relaţia :

$$R^2_{y/x_1,\,x_2} = R^2_{y/x_1} + R^2_{y/x_2} \tag{9.47}$$

Dacă legătura este liniară, atunci se substituie R cu r şi se obţine :

$$R^2_{y/x_1,\,x_2} = r^2_{y/x_1} + r^2_{y/x_2}\,.$$

De unde raportul de corelaţie este :

$$R_{y/x_1,\,x_2} = \sqrt{r^2_{y/x_1} + r^2_{y/x_2}}\,. \tag{9.48}$$

De regulă, însă, în cadrul fenomenelor social-economice factorii sînt interdependenţi între ei şi, de aceea, apare necesitatea luării în considerare a influenţei reciproce a factorilor.

Dacă factorii sînt interdependenţi între ei, atunci $r_{x_1,x} \neq 0$ şi această interinfluenţă, regăsindu-se în valoarea coeficientului de corelaţie multiplă, trebuie să fie eliminată.

Raportul de corelaţie liniară multiplă se calculează atunci folosindu-se coeficienţii de corelaţie simplă, după formula :

$$R_{x_1x_2y} = \sqrt{\frac{r^2_{x_1y} + r^2_{x_2y} - 2r_{x_1y}\,r_{x_2y}\,r_{x_1x_2}}{1 - r^2_{x_1x_2}}} \tag{9.49}$$

Scanned with OKEN Scanner

Această formulă de calcul se foloseşte, în special, atunci cînd au fost deja analizate corelaţiile simple şi, pe baza lor, se trece la măsurarea gradului de intensitate a corelaţiei multiple. De asemenea, se foloseşte şi pentru verificarea liniarităţii funcţiei de estimare.

Dacă nu s-a calculat în prealabil coeficienţii de corelaţie liniară, atunci ei se pot determina folosind formula de calcul simplificat:

$$r_{xy} = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}.$$

Este avantajos să se folosească această formulă, deoarece ea se aplică cu uşurinţă pentru tabelele de corelaţie sau în cazul în care, în prealabil, s-au folosit tabele de calcul al ecuaţiilor de regresie liniară simplă.

În exemplul luat, toate elementele sînt deja calculate pentru determinarea parametrilor funcţiei de estimare (vezi tabelul 9.16) şi se poate trece în continuare pentru calculul celor trei coeficienţi simpli:

a) corelaţia dintre dinamica înzestrării tehnice a muncii şi dinamica producţiei globale industriale pe perioada 1970—1976:

$$r_{x_1y} = \frac{n\Sigma x_1 y - \Sigma x_1 \Sigma y}{\sqrt{[n\Sigma x_1^2 - (\Sigma x_1)^2][n\Sigma y^2 - (\Sigma y)^2]}} =$$

$$= \frac{(7\ 323\ 146) - (1\ 271 \cdot 1\ 730)}{\sqrt{(7 \cdot 234\ 581 - 1\ 271^2)(7 \cdot 449\ 354 - 1\ 730^2)}} = 0{,}994;$$

b) corelaţia dintre dinamica productivităţii muncii şi dinamica produsului global al industriei, pe perioada 1970—1976:

$$r_{x_2y} = \frac{n\Sigma x_2 y - \Sigma x_2 \Sigma y}{\sqrt{[n\Sigma x_2^2 - (\Sigma x_2)^2][n\Sigma y^2 - (\Sigma y)^2]}} =$$

$$= \frac{7 \cdot 306\ 001 - 1\ 202 \cdot 1\ 730}{\sqrt{(7 \cdot 210\ 072 - 1\ 202^2)(7 \cdot 449\ 354 - 1\ 730^2)}} = 0{,}998;$$

c) corelaţia dintre dinamica înzestrării tehnice a muncii şi dinamica productivităţii muncii pe perioada 1970—1976:

$$r_{x_1x_2} = \frac{n\Sigma x_1 x_2 - \Sigma x_1 \Sigma x_2}{\sqrt{[n\Sigma x_1^2 - (\Sigma x_1)^2][n\Sigma x_2^2 - (\Sigma x_2)^2]}} =$$

$$= \frac{7 \cdot 221\ 966 - 1\ 271 \cdot 1\ 202}{\sqrt{(7 \cdot 234\ 581 - 1\ 271^2)(7 \cdot 210\ 072 - 1\ 202^2)}} = 0{,}9925.$$

Comparînd cei trei coeficienţi de corelaţie liniară simplă se constată că între cele trei variabile statistice a existat o strînsă interdependenţă, fie că sînt luate împreună, fie separat două cîte două.

Scanned with OKEN Scanner

Avînd cei trei coeficienţi de corelaţie liniară simplă se poate calcula coeficientul de corelaţie liniară multiplă:

$$R_{x_1x_1y} = \sqrt{\frac{0{,}994^2 + 0{,}998^2 - 2 \cdot 0{,}994 \cdot 0{,}998 \cdot 0{,}9925}{1 - 0{,}9925^2}} = 0{,}999.$$

Şi coeficientul de corelaţie liniară multiplă dedus din coeficienţii de corelaţie simplă confirmă acelaşi grad mare de legătură, dintre cele trei caracteristici, stabilit pe baza ecuaţiilor de regresie ($R_{x_1 \cdot x_2, y} = 0{,}999$).

În exemplul analizat, între cele două caracteristici factoriale există o relaţie de interdependenţă. Dacă cele două caracteristici factoriale sînt independente între ele, atunci coeficientul de corelaţie liniară simplă (r_{x_1, x_2}) este egal cu zero.

În cazul în care caracteristicile factoriale sînt independente între ele, înseamnă că şi coeficientul de regresie din ecuaţia funcţiei liniare simple ce exprimă legătura dintre x_1 şi x_2 este egal cu zero; iar coeficienţii de regresie calculaţi pentru corelaţia simplă dintre fiecare caracteristică factorială şi cea rezultativă sînt egali cu cei corespunzători din ecuaţia funcţiei de regresie multiplă.

De regulă, însă, factorii sînt interdependenţi între ei şi cu cît interdependenţa este mai pronunţată, cu atît vor apărea abateri între coeficienţii de regresie din ecuaţia de corelaţie multiplă şi cei corespunzători din ecuaţia corelaţiei simple.

Pentru exemplificare se vor calcula cele două ecuaţii de regresie simplă ale estimării dinamicii producţiei în funcţie de un singur factor.

a) Ecuaţia medie de tendinţă a dinamicii produsului global al industriei pe perioada 1970—1976 în funcţie de înzestrarea tehnică se va determina cu ajutorul următorului sistem de ecuaţii:

$$na + b\Sigma x_1 = \Sigma y;$$

$$a\ \Sigma x_1 + b\ \Sigma x_1^2 = \Sigma x_1 y.$$

Folosind datele din tabelul 9.16, se obţine:

$$7\ a + 1\,271\ b = 1\,730;$$

$$1271\ a + 234\,581\ b = 323\,146,$$

de unde:

$$a_{x_1} = -183{,}8;$$

$$b_{x_1} = 2{,}37.$$

Deci:

$$\bar{Y}_{x_1} = -183{,}8 + 2{,}37\ x_1.$$

Scanned with OKEN Scanner

b) Ecuaţia medie de tendinţă a dinamicii produsului global în funcţie de dinamica productivităţii muncii, pe perioada 1970—1976, se determină folosind sistemul de ecuaţii :

$$na + b\Sigma x_2 = \Sigma y;$$

$$a\Sigma x_2 + b\Sigma x_2^2 = \Sigma x_2 y.$$

Folosind datele din acelaşi tabel de calcul :

$$7\ a + 1\,202\ b = 1\,730;$$

$$1\,202\ a + 210\,072\ b = 306\,011,$$

de unde :

$$a_{x_2} = -170{,}8;$$

$$b_{x_2} = 2{,}43.$$

Deci :

$$\bar{Y}_{x_2} = -170{,}8 + 2{,}43\ x_2.$$

Comparînd cu ecuaţia de regresie liniară multiplă pentru aceiaşi factori :

$$\bar{Y}_{x_1,\ x_2} = -202{,}38 + 0{,}27\ x_1 + 2{,}35\ x_2$$

se constată că între cei doi factori există o relaţie de interdependenţă, deoarece coeficienţii de regresie pentru acelaşi factor diferă de la corelaţia simplă la cea multiplă. Cum este şi firesc, în corelaţia multiplă cel de-al doilea factor — productivitatea muncii — preia o parte din influenţa primului factor — înzestrarea tehnică a muncii.

9.3.3.3. CORELAŢIA PARŢIALĂ

Cei doi factori (x_1, x_2) fiind interdependenţi între ei, se vor calcula şi indicatori care să măsoare corelaţia parţială.

La calcularea coeficientului de corelaţie multiplă s-a considerat caracteristica rezultativă în funcţie de variaţia concomitentă a celor doi factori, iar în cazul corelaţiei simple s-a analizat legătura separată dintre aceste caracteristici luate cîte două. În continuare se pot analiza şi raporturile de corelaţie parţială, pornind de la corelaţia multiplă, dar măsurînd numai dependenţa dintre aceste caracteristici prin excluderea succesivă a influenţei celuilalt factor. Şi aici se pot folosi relaţiile dintre dispersiile calculate utilizînd ca funcţii de ajustare linia de regresie simplă (Y_{x_1} ; Y_{x_2}) şi linia de regresie multiplă ($Y_{x_1x_2}$) :

$$r_{yx_1 \cdot x_2} = \sqrt{\frac{\sigma^2_{yx_2} - \sigma^2_{y \cdot x_1x_2}}{\sigma^2_{yx_2}}} = \sqrt{1 - \frac{\sigma^2_{yx_1x_2}}{\sigma^2_{yx_2}}} \qquad (9.50)$$

Scanned with OKEN Scanner

şi

$$r_{yx_2 \cdot x_1} = \sqrt{\frac{\sigma^2_{yx_1} - \sigma^2_{y \cdot x_1 x_2}}{\sigma^2_{yx_1}}} = \sqrt{1 - \frac{\sigma^2_{yx_1x_2}}{\sigma^2_{yx_1}}} \qquad (9.51)$$

Coeficienţii de corelaţie parţială se mai pot calcula şi pe baza coeficienţilor simpli, astfel:

$$r_{x_1yx_2} = \frac{r_{x_1y} - r_{x_2y}\, r_{x_1x_2}}{\sqrt{(1 - r^2_{x_2y})(1 - r^2_{x_1x_2})}}; \qquad (9.52)$$

— pentru exemplul analizat, valoarea lui va fi:

$$r_{x_1y \cdot x_2} = \frac{0{,}994 - (0{,}998 \cdot 0{,}9925)}{\sqrt{(1 - 0{,}998^2)(1 - 0{,}9925^2)}} = 0{,}465$$

şi

$$r_{x_2y \cdot x_1} = \frac{r_{x_2y} - r_{x_1y} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{x_1y})(1 - r^2_{x_1x_2})}} \qquad (9.53)$$

care, în cazul respectiv, este:

$$r_{x_2y \cdot x_1} = \frac{0{,}998 - (0{,}994 \cdot 0{,}9925)}{\sqrt{(1 - 0{,}99^{42})(1 - 0{,}9925^2)}} = 0{,}535.$$

Prin coeficienţii de corelaţie parţială se confirmă aceeaşi tendinţă de interacţiune dintre factori şi caracteristica rezultativă, precum şi aceeaşi ierarhizare a factorilor. Primul coeficient de corelaţie parţială arată că în condiţiile în care productivitatea muncii s-ar fi menţinut la un nivel constant, gradul de influenţă asupra dinamicii produsului global ar fi fost de numai 0,465; cel de-al doilea coeficient de corelaţie parţială arată o legătură mai strînsă, ceea ce confirmă că în procesul de creştere a productivităţii muncii au acţionat pozitiv şi alţi factori, influenţa lor regăsindu-se, în final, în dinamica mai rapidă a produsului global faţă de aceea a productivităţii muncii din industrie.

În exemplul analizat s-a luat în considerare calculul raportului de corelaţie pentru serii statistice simple. În cazul distribuţiilor de frecvenţe trebuie să se ţină seama şi de ponderea cu care fiecare valoare individuală intervine în formarea raportului de corelaţie. În acest caz, sistemul de ecuaţii normale pentru aflarea parametrilor $a_0, a_1, \ldots, a_n$ se obţine pe baza produselor de frecvenţă corespunzătoare.

Pentru cazul a două caracteristici factoriale, a cărui raport de corelaţie se formează pe baza unor distribuţii de frecvenţe, sistemul de ecuaţii normale va fi:

$$a_0 \Sigma f + a_1 \Sigma x_1 f_{x_1} + a_2 \Sigma x_2 f_{x_2} = \Sigma y f_y;$$

$$a_0 \Sigma x_1 f_{x_1} + a_1 \Sigma x_1^2 f_{x_1} + a_2 \Sigma x_1 x_2 f_{x_1, x_2} = \Sigma x_1 y f_{x_1 y}$$

$$a_0 \Sigma x_2 f_{x_2} + a_1 \Sigma x_1 x_1 f_{x_1, x_2} + a_2 \Sigma x_2^2 f_{x_2} = \Sigma x_2 y f_{x_2, y}$$

Scanned with OKEN Scanner

iar la raportul de corelaţie:

$$R_{x_1,x_2y} = \sqrt{1 - \frac{\Sigma(y - Y_{x_1,x_2,x_3,\ldots,x_n})^2 f_{x_1,x_2,x_3,\ldots,x_n}}{\Sigma(y - \bar{y})^2 f_y}}. \qquad (9.54)$$

9.3.4. Metode neparametrice de măsurare a legăturilor dintre fenomenele social-economice

După cum s-a arătat în paragrafele precedente, metodele analitice de calcul al corelaţiilor se utilizează în cazul în care există posibilitatea de a se determina forma de manifestare a legăturii, verificată pentru un număr suficient de date care tind să se distribuie normal sau asimptotic normal.

Există, însă, numeroase cazuri cînd distribuţia caracteristicilor nu este normală şi nici nu există informaţii despre parametrii funcţiilor studiate. În aceste situaţii nu se mai pot întrebuinţa formulele indicatorilor analitici de corelaţie, ci trebuie să se folosească alte metode pentru a putea determina existenţa, direcţia şi intensitatea anumitor legături ce se stabilesc între două sau mai multe caracteristici. Aceste metode trebuie să elimine ipoteza privind tipul curbei de distribuţie şi, în consecinţă, să dea posibilitatea unor estimări la cele mai variate tipuri de distribuţie.

Metodele prin care se rezolvă aceste probleme sînt cunoscute sub denumirea de metode neparametrice. Aceste metode, pe lîngă faptul că pot stabili intensitatea unei legături făcînd abstracţie de tipul de distribuţie, permit, de asemenea, măsurarea intensităţii legăturilor nu numai pentru caracteristicile cantitative, dar şi pentru caracteristicile calitative, deoarece în cazul metodelor neparametrice nu se lucrează cu valorile caracteristicilor, ci cu anumite simboluri sau cu un număr de ordine numit rang.

Dintre metodele neparametrice folosite cel mai frecvent sînt: coeficientul de asociere şi coeficienţii de corelaţie a rangurilor a lui Spearman şi Kendall.

9.3.4.1. TABELUL DE ASOCIERE ŞI COEFICIENTUL DE ASOCIERE

Una din metodele cele mai simple ale corelaţiei neparametrice o constituie coeficientul de asociere calculat pe baza tabelului de asociere, care permite — dealtfel ca şi celelalte metode neparametrice — analiza nu numai a caracteristicilor exprimate numeric, dar şi a caracteristicilor calitative. Această metodă se utilizează în special, cînd unităţile purtătoare ale caracteristicilor sînt separate în două grupe sau sînt de forma unor caracteristici alternative (de tipul „da—nu").

Tabelul de asociere este format din două rînduri şi două coloane, de forma:

a	b
c	d

Scanned with OKEN Scanner

în care, în capetele rîndurilor şi coloanelor, se trec variantele celor două caracteristici care se supun asociaţiei, iar în interiorul tabelului, în rubricile lui, se trec frecvenţele corespunzătoare. Deci, în tabelul de asociere se consideră variaţia caracteristicii urmărite în raport cu variaţia caracteristicii de grupare, considerată caracteristică factorială. Produsul *ad* arată gradul de realizare a legăturii dintre *x* şi *y*, iar produsul *bc* lipsa legăturii între aceste două caracteristici cercetate.

Pentru stabilirea valorii numerice a coeficientului de asociere, care să indice existenţa şi intensitatea unei legături, formula cea mai utilizată este cea propusă de **Yule**:

$$Q = \frac{ad - bc}{ad + bc}. \tag{9.55}$$

Este evident că în această situaţie apar cazurile:

a) **independenţa de asociere** cînd:

$$\frac{a}{c} = \frac{b}{d}, \quad \text{deci:} \quad ad - bc = 0;$$

b) asociere completă, care se poate prezenta în mai multe variante:

$$\frac{a \mid 0}{0 \mid d} \quad \text{sau} \quad \frac{a \mid b}{0 \mid d} \quad \text{sau} \quad \frac{a \mid b}{c \mid 0} \quad \text{sau} \quad \frac{0 \mid b}{c \mid d}$$

— prima variantă prezintă un caz de asociere completă absolută cînd valoarea coeficientului este 1;

— a doua variantă prezintă un caz de asociere completă cu sens pozitiv, deoarece locul liber se găseşte pe prima bisectoare (coeficientul = 1);

— a treia variantă prezintă un caz de asociere completă absolută cu valoarea coeficientului egală cu −1;

— a patra variantă prezintă un caz de asociere completă cu sens negativ, locul liber găsindu-se pe a doua bisectoare (coeficientul = −1).

c) cînd gradul de asociere este cuprins între 0 şi ±1 se obţine:

$$\frac{a}{c} \neq \frac{b}{d},$$

în care:

$$ad - bc \neq 0$$

Ca orice coeficient de corelaţie şi acest indicator poate să ia valori între −1 şi +1, arătînd nu numai gradul de intensitate al asocierii celor două caracteristici, dar şi sensul ei. Inconvenientul acestei formule este faptul că, în cazul asociaţiei complete, nu se disting cele patru variante prezentate anterior. De aceea, în acest caz, trebuie făcută menţiunea respectivă în funcţie de modul de asociere a caracteristicilor în tabelul de asociaţie.

Scanned with OKEN Scanner

Pentru exemplificare se va folosi gradul de asociere dintre înzestrarea electro-energetică şi productivitatea muncii într-o ramură de producţie folosind gruparea muncitorilor după poziţia pe care o au faţă de nivelul mediu pe ramură.

Tabelul 9.18

Structura muncitorilor după înzestrarea electro-energetică şi productivitatea muncii în anul 1977 (în %)

	Sub nivelul mediu al productivităţii	Peste nivelul mediu al productivităţii	Total
— Sub nivelul mediu al înzestrării electro-energetice a muncii	45,2	4,6	49,8
— Peste nivelul mediu al înzestrării electro-energetice a muncii	18,4	31,8	50,2
	63,6	36,4	100,0

Din interpretarea datelor din tabel reiese că muncitorii din această ramură se repartizau faţă de nivelul mediu aproape egal din punctul de vedere al înzestrării electro-energetice, în timp ce după productivitatea muncii 63,6% se găseau sub nivelul mediu şi numai 36,4% peste acest nivel. Pentru a avea gradul de asociere a celor două variabile, analizate mai întîi independent, se va calcula coeficientul de asociere:

$$Q = \frac{(45{,}2 \cdot 31{,}8) - (18{,}4 - 4{,}6)}{(45{,}2 \cdot 31{,}8) + (18{,}4 \cdot 4{,}6)} = 0{,}888.$$

Coeficientul de asociere arată o valoare destul de mare la care a contribuit, în special, repartizarea muncitorilor din prima grupă de înzestrare (sub nivelul mediu), care s-a asociat în proporţie de 45,2% cu cei care aveau o poziţie similară din punctul de vedere al productivităţii muncii.

Acest coeficient prezintă avantajul că se poate calcula cu multă rapiditate şi se poate folosi şi în cazul cînd datele provin de la unităţi statistice complexe care, în interiorul lor, pot prezenta forme diferite de distribuţie.

9.3.4.2. COEFICIENŢII DE CORELAŢIE A RANGURILOR

După cum s-a arătat, coeficienţii de corelaţie a rangurilor prezintă avantajul că ei pot fi utilizaţi şi în cazul unor distribuţii asimetrice, în cazul unui număr restrîns de unităţi pentru care nu se poate verifica reprezentativitatea datelor parţiale sau în cazul distribuţiilor unor unităţi complexe. Analiza de corelaţie, în acest caz, se bazează pe înlocuirea valorilor empirice cu rangurile corespunzătoare, ceea ce face posibilă comparaţia dintre variabilele supuse interdependenţei statistice. Această metodă neparametrică mai prezintă avantajul că poate cuprinde în analiza raporturilor de dependenţă dintre fenomene şi caracteristici calitative, care, prin

Scanned with OKEN Scanner

natura lor, nu se pot exprima numeric, dar pot fi „ierarhizate" pe baza unui anumit rang.

Pornind de la ipoteza că între cele două serii de ranguri există concordanţă, seria a doua care reprezintă rangurile caracteristicii rezultative ar trebui să se ordoneze şi ca tot crescător, dacă legătura este directă, şi descrescător în raport cu prima serie de ranguri, dacă legătura este inversă. În cazul existenţei legăturii dintre cele două caracteristici, la aceeaşi unitate observată, trebuie să corespundă acelaşi număr de unităţi care au rang mai mare sau mai mic decît ele. În cazul lipsei de legătură între caracteristicile studiate, ordinea de distribuţie a rangurilor celor două caracteristici este diferită.

În mod corespunzător se pot cuprinde în analiză şi distribuţiile paralele a mai multor caracteristici, cu care se pot realiza mai multe combinaţii, stabilind coeficienţii de corelaţie a rangurilor simpli, parţiali şi multipli, urmărind astfel atît procesul complex de formare a indicatorilor sintetici, cît şi interacţiunea dintre factori.

Cele mai frecvente formule de calcul al coeficientului de corelaţie a rangurilor sînt cele ale lui Spearman şi Kendall.

Coeficientul de corelaţie a rangurilor al lui Spearman este o aplicaţie a coeficientului de corelaţie liniară simplă la distribuţiile celor două şiruri de ranguri.

În cazul corelaţiei rangurilor, cele două serii de valori variabile sînt înlocuite cu două şiruri de ranguri de la 1 la n, ale căror medii şi dispersii sînt egale între ele. Ţinînd seama de faptul că cele două şiruri de ranguri formează cîte o progresie aritmetică 1, 2, 3...,n, înseamnă că:

$$\bar{x} = \bar{y} = \frac{n+1}{2}$$

iar

$$\Sigma(x - \bar{x})^2 = \Sigma(y - \bar{y})^2 = n\,\frac{n^2 - 1}{12}$$

Înlocuind în formula coeficientului de corelaţie liniară se obţine formula de calcul al coeficientului de corelaţie a rangurilor lui Spearman:

$$\gamma_r = 1 - \frac{6\Sigma d^2}{n^3 - n}, \qquad (9.56)$$

în care:

d reprezintă diferenţa de rang între caracteristicile corelate;
n — numărul de unităţi cercetate.

Coeficientul de corelaţie a rangurilor al lui Kendall se calculează după formula:

$$\tau = \frac{S}{1/2\, n(n-1)}, \qquad (9.57)$$

în care:

$$S = P + Q$$

Scanned with OKEN Scanner

Mărimea S reprezintă pe total acele ranguri la care nu există concordanță. Pentru a calcula această mărime, se înscriu rangurile variabilei y, condiționate de rangurile variabilei x. Se calculează pe rînd cele două elemente componente P și Q. Dacă există concordanță și legătura este directă, atunci rangurile lui y trebuie să fie egale cu cele ale lui x și din ce în ce mai mari. De exemplu, pentru unitatea care are rangul 5 în funcție de x, are tot rangul 5 și după y; pentru unitățile care urmează se vor întîlni numai ranguri superioare lui 5. Pentru a măsura eventualele abateri de la această tendință, se vor calcula numărul rangurilor mai mari care urmează și care formează componentul mărimii P și numărul rangurilor următoare mai mici decît rangul unităților luate în calcul și care formează mărimea Q.

Prin adunare, Q are întotdeauna valoare negativă și, deci, suma S — denumită și scor — poate avea valoare pozitivă sau negativă, după cum cele două componente Q și P sînt mai mari în valoare absolută, una în comparație cu alta. De aceea și coeficientul Kendall al corelației rangurilor poate lua valori de la -1 la $+1$, indicînd, în felul acesta, nu numai gradul de intensitate a legăturii, dar și direcția ei.

Pentru a analiza legătura dintre ritmul de creștere a productivității muncii și cea a producției globale industriale, pe perioada 1966—1977, se vor folosi coeficienții de corelație a rangurilor. Este mai indicat aici să se folosească metode neparametrice, deoarece se supun corelației mărimi relative a căror bază diferă de la o ramură la alta și, fiind unități complexe, distribuțiile pe întreprinderi în cadrul ramurilor de producție industrială pot să coincidă cu forme de repartiții de frecvențe pe total.

Tabelul 9.19

Ritmurile medii ale dinamicii productivității muncii și ale producției globale industriale în perioada 1966—1977

(1965=100)

Nr. crt.	Ramura	Productivitatea muncii	Prod. globală	Nr. crt.	Ramura	Productivitatea muncii	Prod. globală
A	B	1	2	A	B	1	2
1	Energie electrică și termică	10,2	12,1	10	Celuloză și hîrtie	7,91	11,6
2	Combustibili	4,9	5,7	11	Sticlă, porțelan, faianță	6,5	12,4
3	Metalurgie feroasă	8,7	12,11	12	Textilă	5,8	12,0
4	Metalurgie neferoasă	7,7	10,4	13	Confecții	6,51	13,8
5	Constr. de mașini și prelucr. metalelor	8,4	16,4	14	Pielărie, blănărie, încălțăminte	5,0	9,5
6	Chimie	10,3	18,0	15	Alimentară	3,5	7,5
7	Extracția minereurilor nemetalifere	9,1	13,4	16	Săpunuri și articole de cosmetică	5,6	10,8
8	Materiale de construcții	7,9	12,12	17	Poligrafie	5,4	5,0
9	Exploatarea și prelucr lemnului	5,41	6,6				

Sursa: Anuarul Statistic al R.S. România, D.C.S., 1978, p. 136;195.

Scanned with OKEN Scanner

Pentru a calcula coeficienţii de corelaţie a rangurilor, se înlocuiesc în ordine crescătoare ramurile industriale prin ranguri şi se înscriu ramurile în funcţie de rangul caracteristicii factoriale (vezi tabelul 9.20).

Tabelul 9.20

Calculul coeficienţilor de corelaţie a rangurilor

Nr. crt.	Ramura	Rangul după:		d	d^2	$S = P + Q$
		productiv. muncii	producţia globală			
A	B	1	2	3	4	5
1	Alimentară	1	4	+3	9	13−3=10
2	Combustibili	2	2	0	0	14−1=13
3	Pielărie, blănărie, încălţăminte	3	5	+2	4	12−2=10
4	Poligrafie	4	1	−3	9	13−0=13
5	Exploatarea şi prelucrarea lemnului	5	3	−2	4	12−0=12
6	Săpunuri şi cosmetice	6	7	+1	1	10−1= 9
7	Textile	7	9	+2	4	8−2= 6
8	Sticlă, porţelan, faianţă	8	13	+5	25	4−5=−1
9	Confecţii	9	15	+6	36	2−6=−4
10	Metalurgie neferoasă	10	6	−4	16	7−0= 7
11	Materiale de construcţii	11	12	1	1	3−3= 0
12	Celuloză şi hîrtie	12	8	−4	16	5−0= 5
13	Construcţii de maşini ...	13	16	+3	9	1−3=−2
14	Metalurgie feroasă	14	11	−3	9	3−0= 3
15	Extracţia minereurilor nemetalifere	15	14	−1	1	1−1= 0
16	Energie electrică şi termică	16	10	−6	36	0−1=−1
17	Chimie	17	17	0	0	0−0= 0
	Total	—	—	$\Sigma d=0$	$\Sigma d^2=178$	$S = 88-8 = 80$

Pe baza datelor din tabel se calculează:

— coeficientul de corelaţie a rangurilor Spearman:

$$\gamma_r = 1 - \frac{6 \cdot 178}{17(17^2 - 1)} = 0{,}782;$$

— coeficientul de corelaţie a rangurilor Kendall:

$$\tau = \frac{80}{\frac{1}{2} \cdot 17(17 - 1)} = 0{,}588.$$

Este de remarcat faptul că de obicei coeficientul de corelaţie al rangurilor după formula lui Kendall este mai mic decît cel al lui Spearman. În cazul de faţă s-a obţinut un coeficient al rangurilor Kendall (0,588)

Scanned with OKEN Scanner

mai mic decît al lui Spearman (0,782) dar care indică, ca şi primul, o corelaţie moderată între cele două variabile. Aceasta confirmă faptul că ritmul de creştere a producţiei mai depinde şi de alţi factori, în afara productivităţii muncii.

9.3.5. Analiza raporturilor de corelaţie cu decalaj

În general, în dinamică, între două sau mai multe variabile statistice se realizează o legătură de corelaţie în cadrul aceleiaşi perioade de timp. Dar, în practica statistică se întîlnesc şi alte feluri de legături, a căror interdependenţă se produce după o anumită perioadă de timp.

Este cunoscut faptul că, de regulă, între perioada introducerii unei tehnologii avansate şi creşterea volumului producţiei, ca urmare a creşterii productivităţii muncii, poate să existe un anumit decalaj. O asemenea legătură se poate urmări şi între creşterea retribuţiei reale şi reducerea costului producţiei, care formează principala premisă materială a reducerii preţurilor cu amănuntul, sau chiar direct legătura dintre reducerea costului şi a preţului de vînzare la acelaşi produs.

O tratare amănunţită a acestei forme de corelaţie cu întîrziere este aceea a legăturii între indicii volumului importului de materii prime şi cei ai volumului producţiei industriale.[1]

Pentru a stabili mărimea decalajului (întîrzierii) de unde începe corelaţia celor două serii de indicatori, autorii, în lucrarea citată, introduc în plus notaţiile necesare în aşa fel încît formula coeficientului de corelaţie simplă devine :

$$r_h = \frac{\sum_{h=1}^{n} (x_t - \bar{x})(y_{t-h} - \bar{y})}{(n-h)\,\sigma_x \sigma_y}, \tag{9.58}$$

în care :

t reprezintă numărul de termeni ai seriei iniţiale ;

h — numărul de termeni cu care se decalează cea de-a doua serie.

Deci, între t şi h există relaţia : $t = h + 1 ; h + 2 ; \ldots, n$. Acest coeficient arată corelaţia între termenii unei serii de timp t şi cealaltă $t-h$.

Pentru fixarea decalajului care asigură stabilizarea unei corelaţii optime, autorii recomandă ca fiind indispensabilă reprezentarea grafică a celor două serii dinamice. Mai mult, autorii arată că un mijloc eficace pentru rezolvarea cît mai corectă a acestei probleme este ca cele două curbe să fie făcute separat pe hîrtie de calc, la aceeaşi scară şi glisînd cel de-al doilea grafic, se va putea stabili cu uşurinţă unitatea de timp la care începe corelaţia.

Pentru exemplificare (tab. 9.21) s-au folosit datele cu privire la dinamica investiţiilor totale şi a producţiei globale agricole între anii 1970—1978 într-o întreprindere agricolă de stat (date convenţionale).

---

[1] E. Morice, F. Chartier, *Méthode statistique, Deuxième partie*, Paris, Imprimerie Nationale, 1964, p. 501—508.

Scanned with OKEN Scanner

Interpretînd aceste două serii de date se poate observa că ambele caracteristici prezintă o tendință de creștere, dinamica investițiilor devansînd dinamica producției.

Din datele tabelului se poate observa că, deși investițiile au crescut în primul an cu aproape 30%, producția globală agricolă realizează această creștere abia între anii 1970—1973, cu 2 ani mai tîrziu. Se poate aprecia că, corelația dintre cele două caracteristici, luate în dinamică, începe să se realizeze cu o întîrziere de 2 ani. Deci, ținînd seama de această apreciere vizuală a decalajului dintre cele două serii dinamice, se pot forma cele două serii cronologice t și $t-h$, adică :

$$x_1 ; \; x_2 ; \; x_3, \; x_4 ; \; x_5 ; \; x_6 ;$$

$$y_3 ; \; y_4 ; \; y_5 ; \; y_6 ; \; y_7 ; \; y_8.$$

Deci, în exemplul considerat, prima serie de timp t — dinamica investițiilor — va include indicii cu bază fixă, cuprinși în perioada 1971—1976, iar pentru cea de-a doua serie de timp $t-h$ — privind dinamica producției globale agricole, indicii cu bază fixă cuprinși între anii 1973—1978.

Tabelul 9.21

Dinamica investițiilor și a producției globale agricole între anii 1970—1978.

(1970 = 100)

Anii	Valoarea investițiilor	Producția globală agricolă
0	1	2
1971	129	112
1972	123	124
1973	142	135
1974	177	142
1975	198	146
1976	199	150
1977	192	156
1978	256	168

În acest fel, se consideră dinamica producției globale agricole pe perioada 1973—1978 în funcție de dinamica investițiilor din perioada 1971—1976, deci, dintr-o perioadă de timp corespunzătoare, decalată cu un avans de doi ani (tabelul 9.22).

$$\bar{x} = \frac{\Sigma x}{n} = \frac{968}{6} = 161{,}3 \; ; \; \sigma_x = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}} = \sqrt{\frac{5\,897{,}74}{6}} = 31{,}3 \; ;$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{897}{6} = 149 \; ; \quad \sigma_y = \sqrt{\frac{\Sigma(y - \bar{y})^2}{n}} = \sqrt{\frac{663{,}50}{6}} = 10{,}5.$$

Scanned with OKEN Scanner

Tabelul 9.22

Corelația cu decalaj între dinamica investițiilor și dinamica producției globale agricole în perioada 1970—1978

(1970 = 100)

Nr. crt.	x_t	y_{t-h}	$x_t - \bar{x}$	$(x_t - \bar{x})^2$	$y_{t-h} - \bar{y}$	$(y_{t-h} - \bar{y})^2$	$(y_t - \bar{x})(y_{t-h} - \bar{y})$
0	1	2	3	4	5	6	7
1	129	135	−32,3	1 043,29	−14,5	210,25	+468,35
2	123	142	−38,3	1 466,89	− 7,5	56,25	+287,25
3	142	146	−19,3	372,49	− 3,5	12,25	+ 67,55
4	177	150	−15,7	246,49	+ 0,5	0,25	+ 7,85
5	198	156	36,7	1 346,89	+ 6,5	42,25	+238,55
6	199	168	37,7	1 421,29	+18,5	342,25	+697,45
	968	897	—	5 897,34	—	663,50	1 767,00

$$r_{x,y} = \frac{1\,767,00}{6 \cdot 31,3 \cdot 10,5} = \frac{1\,767,00}{1\,791,90} = 0,89609.$$

Deci, între dinamica investițiilor și producția agricolă globală, în perioada 1970—1978, în întreprinderea agricolă de stat, s-a realizat o corelație cu un decalaj de doi ani, destul de ridicată (0,89)[1]. Pe de altă parte analiza raporturilor de corelație cu decalaj a permis să se poată face aprecieri asupra condițiilor create deja pentru creșterea producției globale agricole în etapa care urmează. Mai mult, se poate aprecia că pentru o analiză a raporturilor de corelație ale producției pe o etapă următoare, se poate păstra același decalaj de doi ani.

Ceea ce trebuie subliniat aici este necesitatea luării în considerare a unui număr suficient de mare de termeni pentru cele două serii inițiale de date, pentru ca într-adevăr să se poată desprinde cît mai corect care este tendința de dezvoltare și de formare a legăturilor dinamice de corelație. Numai în acest fel se poate urmări, pe etape, procesul obiectiv de realizare a interdependenței dintre ele. În acest fel metodele și procedeele utilizate în cercetarea statistică a legăturilor de corelație ce se formează în dinamică sînt numai instrumente de calcul; determinantă pentru interpretarea și măsurarea legăturii este analiza teoretică calitativă care precede calculul statistic.

Dacă, însă, pe grafic rezultă o corelație neliniară, atunci se calculează numai raportul de corelație corespunzător funcției analitice ce exprimă legătura cu decalaj dintre caracteristica (caracteristicile) factorială(le) și cea rezultativă.

În concluzie, alegerea modelelor de calcul și interpretare a corelațiilor dintre fenomenele social-economice trebuie să se facă cu mult discernămînt, astfel încît să se poată obține rezultatele veridice pentru ilustrarea numerică a fenomenelor și proceselor sociale de masă care au fost studiate.

[1] Pentru comparație se poate calcula și coeficientul de corelație liniară simplă pentru distribuția bidimensionala concomitentă. Folosind datele din tabelul 9.21, se obține un coeficient de corelație liniară simplă de 0,573, ceea ce confirmă necesitatea folosirii în cazul dat a corelației cu decalaj și nu a corelației directe dintre cele două serii cronologice.

Scanned with OKEN Scanner

CAPITOLUL X

ANALIZA SERIILOR CRONOLOGICE

Statistica social-economică studiază fenomenele şi procesele sociale în continua lor dezvoltare, în vederea cunoaşterii modului concret de transformare a acestora în fiecare etapă dată. În munca de analiză concretă a dezvoltării fenomenelor în timp, statistica foloseşte ca instrument principal de cercetare indicatorii obţinuţi din prelucrarea statistică a seriilor cronologice.

10.1. NOŢIUNI. PARTICULARITĂŢI

Seria cronologică este formată din două şiruri de date paralele, în care primul şir arată variaţia caracteristicii de timp, iar cel de-al doilea şir variaţia fenomenului sau caracteristicii cercetate, de la o unitate de timp la alta. După cum s-a mai arătat, seriile cronologice se mai numesc şi serii de timp sau serii dinamice.

La analiza seriilor cronologice trebuie avut în vedere o serie de proprietăţi ale acestora. Ele se caracterizează prin : *variabilitatea*, *omogenitatea*, *periodicitatea* şi *interdependenţa* termenilor prezentaţi.

Variabilitatea termenilor unei serii cronologice apare ca urmare a faptului că fiecare termen se obţine prin centralizarea unor date individuale diferite ca nivel de dezvoltare. Existenţa unor date individuale diferite se explică prin faptul că, în cadrul fenomenelor sociale acţionează, pe lîngă cauzele esenţiale, determinante şi un număr suficient de mare de cauze neesenţiale, a căror mod de asociere se poate schimba de la o perioadă la alta.

Este cunoscut faptul că, cu cît acţiunea factorilor întîmplători este mai puternică, cu atît variaţia în cadrul colectivităţii este mai mare, iar influenţa lor se imprimă în mai mare măsură asupra modului de manifestare a fenomenelor în fiecare etapă.

În plus, la analiza seriilor dinamice trebuie avut în vedere şi faptul că ele se întocmesc pentru unităţi complexe, la nivelul cărora gradul de variaţie al indicatorilor este mult mai mare decît la nivelul unităţilor simple, cuprinzînd şi variaţiile de structură de la o unitate de timp la alta.

Avînd în vedere această trăsătură, este necesar ca, analizînd o serie cronologică, să se măsoare atît gradul şi forma de influenţă a factorilor esenţiali, cît şi gradul de abatere de la tendinţa generală rezultată din influenţa factorilor neesenţiali, cu caracter întîmplător.

Omogenitatea termenilor trebuie înţeleasă în sensul că în aceeaşi serie nu pot fi înscrise decît fenomene de acelaşi gen, care sînt rezultatul acţiunii aceloraşi legi.

Aceasta impune ca, pentru fiecare cercetare în parte, să se verifice în ce măsură datele sînt omogene între ele şi, deci, se pot înscrie în aceeaşi serie statistică.

Scanned with OKEN Scanner

Pentru a asigura omogenitatea datelor statistice cuprinse într-o serie dinamică, trebuie să se ţină seama de rezolvarea unor probleme legate atît de forma de exprimare, cît şi de conţinutul indicatorilor. Astfel, ţinînd seama de primul aspect va trebui să se aibă în vedere ca indicatorii cuprinşi în aceeaşi serie cronologică să fie exprimaţi în aceleaşi unităţi de măsură, iar pentru indicatorii calculaţi valoric să se ţină seama şi de modificările de preţuri intervenite de la o perioadă la alta.

Fără îndoială, însă, că problema omogenităţii indicatorilor unei serii dinamice se rezolvă numai în măsura în care se ţine seama de cel de-al doilea aspect, respectiv se asigură acestora un conţinut asemănător. În practica statistică această problemă se rezolvă prin compararea datelor ce urmează să se înscrie în aceeaşi serie, atît din punctul de vedere al culegerii, cît şi al prelucrării informaţiei statistice.

O altă trăsătură caracteristică a seriilor dinamice o constituie *periodicitatea termenilor* din care este formată seria. După cum se ştie, fiecare fenomen este legat în mod obiectiv de anumite condiţii specifice de timp şi de spaţiu.

Alegerea unităţii de timp la care se referă datele unei serii cronologice trebuie făcută în raport cu scopul cercetării, al conţinutului şi al posibilităţilor de măsurare a fiecărui indicator. De exemplu, producţia industrială se poate urmări atît în unităţi de timp mai mici (ziua, decada, luna), cît şi în unităţi mai mari de timp (trimestrul, semestrul, anul), în timp ce producţia agricolă totală nu se poate urmări decît pe ani întregi (agricoli sau calendaristici). Pe elemente de calcul, producţia agricolă se poate urmări şi pe perioade mai scurte. Se pot întîlni, de asemenea, cazuri cînd unele caracteristici sînt influenţate în variaţia lor de schimbarea anotimpurilor, cu alte cuvinte, apar fenomene cu caracter sezonier (lunar sau trimestrial).

Această formă de manifestare a fenomenelor în timp necesită studierea periodicităţii fenomenelor şi alcătuirea seriilor dinamice în funcţie de această particularitate.

În legătură cu caracteristica de timp, se pune nu numai problema alegerii unităţilor de timp la care se referă fiecare indicator, ci şi alegerea întregii etape pentru care se prezintă datele. Aceasta înseamnă că, înainte de a trece la întocmirea seriei, să se facă o analiză a condiţiilor în care s-a dezvoltat fenomenul cercetat şi care să permită cunoaşterea etapei în cadrul căreia se pot urmări modificările intervenite în evoluţia sa. În acest sens, deosebit de importantă este alegerea primului an — anul de bază — deoarece mărimea indicatorilor obţinuţi din prelucrare va depinde tocmai de nivelul acestui an. În ţara noastră, ca ani reprezentativi pentru seriile cronologice se consideră anii : 1938, 1950, 1960, 1965 ,1970, 1975, ani care marchează anumite etape în dezvoltarea ascendentă a economiei noastre naţionale.

În cazul unei întreprinderi, pe lîngă aceste serii de date este necesar să se analizeze şi seriile formate pe baza anilor în care s-au produs modificări în ceea ce priveşte structura organizatorică a procesului de producţie şi a procesului de muncă şi, în funcţie de această perioadă, să se analizeze dinamica diferiţilor indicatori.

Interdependenţa termenilor unei serii cronologice apare ca urmare a respectării principiului unităţii de timp şi spaţiu : indicatorii prezentaţi

Scanned with OKEN Scanner

sînt valori succesive ale aceloraşi fenomene înregistrate la nivelul aceleiaşi unităţi teritoriale sau administrative. Aceasta face ca valoarea fiecărui indicator să depindă într-o oarecare măsură de valoarea indicatorului precedent, reflectînd faptul că fenomenele sociale şi economice sînt rezultatul unor legi obiective, ce se manifestă sub formă de tendinţă şi care poate fi urmărită pe o perioadă mare de timp.

De aceea, în cazul seriilor cronologice — dată fiind interdependenţa dintre termeni — se pune problema cunoaşterii liniei (curbei) de tendinţă specifică fiecărei etape de dezvoltare şi care, în sens statistic, exprimă într-o formă cantitativă însăşi acţiunea legii care le determină.

Luînd în considerare toate aceste particularităţi, analiza statistică a seriilor cronologice trebuie să se bazeze pe un sistem de indicatori care să caracterizeze multiplele relaţii cantitative din interiorul seriei şi pe toată perioada la care se referă datele.

10.2. FELURILE SERIILOR CRONOLOGICE

În prezentarea dinamică a fenomenelor se pot întîlni mai multe feluri de serii dinamice. Clasificarea seriilor cronologice se face în funcţie de modul de exprimare a indicatorilor, după modul de exprimare a timpului la care se referă datele.

În funcţie de modul de exprimare a indicatorilor din care este formată seria, seriile cronologice pot fi:

— serii cronologice formate din indicatori absoluţi;
— serii cronologice formate din indicatori relativi;
— serii cronologice formate din indicatori medii.

Seriile cronologice formate din indicatori absoluţi reprezintă forma de bază a seriilor dinamice. Pe baza datelor unei serii dinamice formată din indicatori absoluţi se pot obţine indicatorii generalizatori pe întreaga perioadă. Aceste serii statistice se obţin prin operaţia de concentrare a datelor statistice pentru fiecare unitate de timp.

Seriile cronologice formate din indicatori relativi constituie un mijloc de reprezentare a unor mărimi derivate, calculate fie sub formă de mărimi relative de dinamică, fie sub formă de mărimi relative de coordonare, fie sub formă de mărimi relative de structură. Acestea se pot exprima fie în unităţi de măsură concrete, fie prin numere abstracte, de regulă, sub formă de procente. Indiferent de forma de exprimare a indicatorilor din care este formată seria, în cazul mărimilor relative este obligatoriu ca în titlul sau în afara tabelului să se specifice care este baza de raportare, pentru ca interpretarea datelor să se facă corect. De exemplu, dacă sînt mărimi relative de dinamică trebuie specificat ce perioadă se consideră ca bază de raportare sau, dacă sînt mărimi de structură, care este volumul colectivităţii faţă de care s-au obţinut greutăţile specifice prezentate.

Seriile cronologice formate din mărimi medii se folosesc ca mijloc de prezentare a evoluţiei unor caracteristici calitative ce apar sub formă de categorii medii: productivitatea muncii; randamentul mediu; recolta medie la hectar; retribuţia medie etc. De asemenea, se folosesc astfel de serii şi pentru unele caracteristici cantitative atunci cînd ele trebuie incluse în analiza unor fenomene ce se produc în cadrul unui interval de timp,

Scanned with OKEN Scanner

ca de exemplu : valoarea medie anuală a fondurilor fixe ; număr mediu al personalului muncitor etc. Caracteristic pentru aceste serii este faptul că ele se pot supune în continuare unei prelucrări statistice, în vederea obţinerii şi a altor indicatori derivaţi, care să permită caracterizarea formei lor de evoluţie, la fel ca şi cele formate din indicatorii absoluţi.

Seriile cronologice se mai pot clasifica în funcţie de timpul la care se referă datele :

— serii cronologice de intervale (perioade) de timp ;
— serii cronologice de momente.

Seriile cronologice de intervale (perioade) de timp sînt seriile statistice în care fiecare indicator reprezintă rezultatul unui proces social-economic pe fiecare perioadă de timp folosită în prezentarea datelor.

Astfel de serii se pot întîlni în prezentarea evoluţiei producţiei şi circulaţiei de mărfuri, a mărimii investiţiilor, a numărului de călători transportaţi cu mijloace de transport în comun, date cu privire la mişcarea naturală a populaţiei etc. Ele se întocmesc pentru indicatorii însumabili pe o anumită perioadă de timp care determină periodicitatea cu care se prezintă termenii seriei. Termenii seriei de intervale pot fi cumulaţi obţinîndu-se un indicator totalizator pe întreaga serie. De asemenea, şi periodicitatea înscrierii termenilor unei serii de intervale se poate modifica. De exemplu, prezentînd datele pe trimestrele anului pe un număr de cinci ani, cumulînd cîte patru termeni de la o periodicitate trimestrială, se trece la o periodicitate anuală, reducînd numărul termenilor, de la 20 în prima serie, la 5 în cea de a doua.

Seriile cronologice (dinamice) de momente sînt acelea în care fiecare indicator caracterizează mărimea la care a ajuns caracteristica urmărită sau volumul colectivităţii în momentul de calcul ; de exemplu, puterea instalată, exprimată în mii Kw, la sfîrşitul fiecărui an. Aceste serii se întocmesc pentru variabile statistice însumabile în orice moment. De aceea se pot întîlni serii de momente cu intervale egale sau neegale între termenii prezentaţi.

Pentru seria de momente este caracteristic faptul că termenii ei nu se pot cumula în vederea obţinerii unui indicator statistic totalizator pe întreaga perioadă, ca în cazul seriilor de intervale. Caracterizarea nivelului atins pe întreaga perioadă nu se poate face, în acest caz, decît pe baza unui indicator mediu.

Scopul analizei datelor unei serii cronologice poate fi acela de a caracteriza modul de dezvoltare a fenomenelor sociale şi economice pe o perioadă expirată, în vederea extrapolării datelor statistice pentru fundamentarea diferitelor calcule de prognoză. Pentru realizarea acestui scop este necesar să se rezolve următoarele probleme cu ajutorul cărora se poate interpreta statistic dinamica fenomenelor :

— prelucrarea datelor unei serii dinamice în vederea obţinerii sistemului de indicatori statistici, sub forma de mărimi absolute, relative sau medii, care caracterizează în ansamblu seria sau relaţiile din interiorul ei ;

— determinarea tendinţei de lungă şi scurtă durată, ce se obţine prin aplicarea procedeelor de ajustare a termenilor unei serii cronologice în funcţie de timp ;

Scanned with OKEN Scanner

— determinarea gradului de influenţă a valului sezonier pentru cazul în care fenomenele supuse analizei statistice sînt afectate de schimbarea anotimpurilor ;

— determinarea valorilor probabile pentru etapa ce urmează, prin extrapolarea seriilor de date statistice care au fost supuse prelucrării.

10.3. PRELUCRAREA SERIILOR CRONOLOGICE DE INTERVALE

Seriile cronologice — după cum s-a arătat — prezintă, de regulă, indicatori cu variaţii mari de la o unitate de timp la alta. Pentru a putea caracteriza statistic modul de dezvoltare a fenomenelor este necesar ca datele prezentate în funcţie de timp să fie supuse prelucrării. Sînt supuse unei astfel de prelucrări, în special, seriile dinamice formate din indicatori absoluţi, mărimi relative de intensitate sau mărimi medii. În urma prelucrării statistice a seriilor dinamice se obţin indicatori absoluţi, relativi şi medii, care, împreună, permit caracterizarea sub raport statistic a dezvoltării fenomenelor studiate prin interpretarea tendinţei obiective de dezvoltare a acestora în fiecare etapă dată.

Indicatorii absoluţi, care caracterizează o serie cronologică, sînt :

y_i — nivelurile absolute ale termenilor seriei ;

$\Delta_{i/0}$ — modificarea absolută (spor sau scădere absolută) calculată cu bază fixă ;

$\Delta_{i/i-1}$ — modificarea absolută (spor sau scădere absolută) calculată cu bază în lanţ.

Indicatorii relativi, care caracterizează o serie cronologică, sînt :

$I_{i/0}$ — indicele de dinamică calculat cu bază fixă ;

$I_{i/i-1}$ — indicele de dinamică calculat cu bază în lanţ ;

$R_{i/0}$ — ritmul de creştere (scădere) întîlnit în literatura de specialitate şi sub denumirea de indice al ritmului sporului, calculat cu bază fixă ;

$R_{i/i-1}$ — ritmul de creştere (scădere) calculat cu bază în lanţ ;

$A_{i/0}$ — valoarea absolută a unui procent de creştere (scădere) cu baza fixă ;

$A_{i/i-1}$ — valoarea absolută a unui procent de creştere (scădere) cu bază în lanţ.

Indicatorii medii, care caracterizează o serie cronologică, sînt :

$\bar{y}$ — nivelul mediu al unei serii cronologice de intervale ;

$\bar{\Delta}$ — nivelul mediu al sporului (scăderii) absolute ;

$\bar{I}$ — indicele mediu al dinamicii ;

$\bar{R}$ — ritmul mediu de creştere.

Scanned with OKEN Scanner

Calculul sistemului de indicatori se face diferenţiat pentru o serie de intervale şi pentru o serie de momente. La seria de momente nivelul mediu se calculează prin media cronologică ($\bar{y}_{cr}$).

10.3.1. Indicatorii absoluţi care caracterizează seriile cronologice

Indicatorii absoluţi ai unei serii cronologice se exprimă în unităţile concrete de măsură în care se măsoară şi fenomenul supus cercetării.

Pentru ilustrarea modului de calcul al acestor indicatori se ia exemplul din tabelul 10.1.

Tabelul 10.1

Producţia de energie electrică a R.S.R. în anii 1965—1975

Anii	Producţia (în mil—Kwh) (y)
1965	17 215
1966	20 806
1967	24 769
1968	27 828
1969	31 509
1970	35 088
1971	39 454
1972	43 439
1973	46 779
1974	49 063
1975	53 721

Sursa: Anuarul Statistic al R.S.R., Bucureşti, D.C.S., 1971, p. 198—199; 1976, p. 112—113.

Dacă se notează cu t variabila de timp şi cu y producţia de energie electrică pe unităţi de timp — seria va lua valori de la y_0 (perioada de bază a seriei) la y_n (ultimul termen al seriei), adică :

$$y_0 = 17\,215\,;\ y_1 = 20\,806\,;\ \ldots\ y_{10} = 53\,721.$$

Aceste mărimi constituie însăşi indicatorii de nivel ai seriei cronologice care se studiază. Dacă se notează primul termen cu y_0 şi ultimul cu y_n, înseamnă că seria este formată din $n + 1$ termeni.

Fiind o serie cronologică de intervale cu indicatori exprimaţi în mărimi absolute, se poate trece la prelucrarea seriei prin calcularea indicatorilor absoluţi, relativi şi medii.

Pentru simplificarea prezentării modului de calcul al acestor indicatori şi a urmăririi legăturii dintre indicatorii absoluţi şi cei relativi, se vor trece rezultatele în coloanele tabelului 10.2.

Primii indicatori care s-au calculat aici (vezi coloanele 2 şi 3) sînt modificările absolute, care pot fi interpretate ca *sporuri* (*scăderi*) *absolute* de la o unitate la alta. În teoria şi practica statisticii sporul absolut se mai numeşte şi creşterea absolută sau excedent, iar scăderea absolută poate avea uneori sens de deficit sau de economie absolută. Interpretarea se face în funcţie de conţinutul şi de tendinţa obiectivă de creştere (micşorare) a fenomenului cercetat.

Scanned with OKEN Scanner

Tabelul 10.2

Dinamica producției de energie electrică în R.S.R. între anii 1965—1975

Anii	Prod. în mil. Kwh	Modificări absolute (±) mil. Kwh		Indicii de dinamică %		Ritmul de creștere %	
		cu bază fixă	cu bază în lanț	cu bază fixă	cu bază în lanț	cu bază fixă	cu bază în lanț
	y	$\Delta_{i/0}$	$\Delta_{i/i-1}$	$I_{i/0}$	$I_{i/i-1}$	$R_{i/0}$	$R_{i/i-1}$
A	1	2	3	4	5	6	7
1965	17 215	—	—	100	—	—	—
1966	20 806	3 591	3 591	120,8	120,8	20,8	20,8
1967	24 769	7 554	3 963	143,9	119,0	43,9	19,0
1968	27 828	10 613	3 059	161,6	112,4	61,6	12,4
1969	31 509	14 294	3 681	183,0	113,2	83,0	13,2
1970	35 088	17 873	3 579	203,8	111,4	103,8	11,4
1971	39 454	22 239	4 366	229,2	112,4	129,2	12,4
1972	43 439	26 224	3 985	252,3	110,1	152,3	10,1
1973	46 779	29 564	3 340	271,7	107,7	171,7	7,7
1974	49 063	31 848	2 284	285,0	104,9	185,0	4,9
1975	53 721	36 506	4 658	312,0	109,5	212,0	9,5
—	389 671 $\sum_{i=0}^{n} y_i$	—	36 506 $\sum_{i=1}^{n} \Delta_{i/i-1}$	—	3,12 $\prod_{i=1}^{n} I_{i/i-1}$	—	—

Sporul (scăderea) absolut poate fi calculat fie față de nivelul unei singure perioade considerată ca bază de referință, fie de la o perioadă de timp la alta. În primul caz se obține sporul (scăderea) cu bază fixă, în cel de-al doilea caz se obține sporul (scăderea) cu bază în lanț.

Sporul cu bază fixă (col. 2. tabelul 10.2) se notează cu $\Delta_{i/0}$ și se obține ca diferență între nivelul fiecărei perioade y_i și nivelul perioadei de referință y_0, adică:

$$\Delta_{i/0} = y_i - y_0 \qquad (10.1)$$

De exemplu:

$$\Delta_{1/0} = y_1 - y_0 = 20\,806 - 17\,215 = 3\,591 \text{ mil. Kwh};$$

$$\Delta_{2/0} = y_2 - y_0 = 24\,769 - 17\,215 = 7\,554 \text{ mil. Kwh.}$$

$$\vdots$$

$$\Delta_{10/0} = y_{10} - y_0 = 53\,721 - 17\,215 = 36\,506 \text{ mil. Kwh.}$$

Sporul cu bază în lanț (col. 3 tabelul 10.2) se notează cu $\Delta_{i/i-1}$ și se obține ca diferență între nivelul fiecărui an (y_i) și nivelul anului precedent (y_{i-1}), adică:

$$\Delta_{i/i-1} = y_i - y_{i-1}, \qquad (10.2)$$

Scanned with OKEN Scanner

deci :

$$\Delta_{1/0} = y_1 - y_0 = 20\,806 - 17\,215 = 3\,591 \text{ mil. Kwh};$$

$$\Delta_{2/0} = y_2 - y_1 = 24\,769 - 20\,806 = 3\,963 \text{ mil. Kwh};$$

$$\vdots$$

$$\Delta_{10/9} = y_{10} - y_9 = 53\,721 - 49\,063 = 4\,658 \text{ mil. Kwh.}$$

Între sporurile cu bază în lanţ şi sporurile cu bază fixă există următoarea relaţie : suma sporurilor cu bază în lanţ este egală cu sporul cu bază fixă pe întreaga perioadă care a intrat în calcul :

$$\sum_{i=1}^{n} \Delta_{i/i-1} = \Delta_{n/0} \tag{10.3}$$

ceea ce înseamnă :

$$(y_1 - y_0) + (y_2 - y_1) + \ldots + (y_n - y_{n-1}) = y_n - y_0.$$

În exemplul prezentat va fi :

$$\Delta_{1/0} + \Delta_{2/1} = \Delta_{2/0};\ 3\,591 + 3\,963 = 7\,554;$$

$$\Delta_{1/0} + \Delta_{2/1} + \Delta_{3/2} = \Delta_{3/0};\ 3\,591 + 3\,963 + 3\,059 = 10\,613;$$

$$\vdots$$

$$\Delta_{1/0} + \Delta_{2/1} + \Delta_{3/2} + \ldots + \Delta_{10/9} = \Delta_{10/9};$$

$$3\,591 + 3\,963 + 3\,059 + 3\,681 + 3\,579 + 4\,366 + 3\,985 +$$

$$+ 3\,340 + 2\,284 + 4\,658 = 36\,506.$$

Folosind această relaţie de bază, în mod corespunzător se poate trece de la sporurile cu bază fixă la cele cu bază în lanţ. Astfel, dacă din sporul cu bază fixă a unei perioade se scade sporul cu bază fixă a unei perioade precedente, se obţine sporul cu bază în lanţ corespunzător :

$$\Delta_{i/0} - \Delta_{i-1/0} = \Delta_{i/i-1} \tag{10.4}$$

$$(y_i - y_0) - (y_{i-1} - y_0) = y_i - y_0 - y_{i-1} + y_0 = y_i - y_0,$$

deci :

$$\Delta_{2/0} - \Delta_{1/0} = \Delta_{2/1} = 7\,554 - 3\,591 = 3\,963$$

$$\vdots$$

$$\Delta_{10/0} - \Delta_{9/0} = \Delta_{10/9} = 36\,506 - 31\,848 = 4\,658.$$

Aceste relaţii se folosesc atunci cînd nu se dispune de date absolute, ci se cunosc sau numai sporurile cu bază fixă, sau numai sporurile cu bază în lanţ. În acelaşi timp, cu ajutorul acestor relaţii se poate verifica exactitatea calculelor.

Scanned with OKEN Scanner

10.3.2. Indicatorii relativi ai seriilor cronologice

Indicatorii relativi ai seriilor dinamice ocupă un loc deosebit de important în analiza concretă a dinamicii fenomenelor social-economice. Aceşti indicatori sînt folosiţi pe larg în documentele de partid şi de stat în legătură cu stabilirea proporţiilor şi corelaţiilor între diferitele ramuri şi sectoare ale economiei naţionale. Ritmuri înalte şi diferenţiate pe ramuri, judeţe şi categorii economice imprimă întregii economii o dezvoltare armonioasă, ascendentă.

Mărimile relative folosite în analiza dinamicii fenomenelor se pot referi la două aspecte : să măsoare de cîte ori nivelul dintr-o perioadă este mare decît cel ales ca bază de comparare, sau să măsoare cu cît s-a modificat nivelul fenomenului din perioada raportată faţă de cel din perioada de raportare.

Mărimea relativă, care arată de cîte ori s-a modificat un fenomen în timp, se numeşte *indice de dinamică*[1] şi se poate calcula cu bază fixă şi cu bază în lanţ.

Indicele de dinamică cu bază fixă se notează cu $I_{i/0}$(col. 4, tabelul 10.2) şi se calculează ca raport între nivelul fiecărui an şi nivelul anului ales ca bază; se exprimă, de regulă, sub formă de procente :

$$I_{i/0} = \frac{y_i}{y_0} \cdot 100, \tag{10.5}$$

adică :

$$I_{1/0} = \frac{y_1}{y_0} = \frac{20\,806}{17\,215} = 1{,}208 \text{ sau } 120{,}8\%;$$

$$I_{2/0} = \frac{y_2}{y_0} = \frac{24\,769}{17\,215} = 1{,}439 \text{ sau } 143{,}9\%;$$

$$\vdots$$

$$I_{10/0} = \frac{y_{10}}{y_0} = \frac{53\,721}{17\,215} = 3{,}12 \text{ ori.}$$

Indicele de creştere cu bază în lanţ ($I_{i/i-1}$) se calculează ca raport între nivelul fiecărui an şi nivelul anului precedent (col. 5, tabelul 10.2). Şi acesta se exprimă, de regulă, tot sub formă de procente :

$$I_{i/i-1} = \frac{y_i}{y_{i-1}} \cdot 100, \tag{10.6}$$

[1] Cînd acest indice este supraunitar se mai numeşte şi indice de creştere, iar cînd este subunitar apare ca un indice de scădere sau reducere, după conţinutul indicatorului comparat.

adică :

$$I_{1/0} = \frac{y_1}{y_0} = \frac{20\,806}{17\,215} = 1{,}208 \text{ sau } 120{,}8\%;$$

$$I_{2/1} = \frac{y_2}{y_1} = \frac{24\,769}{20\,806} = 1{,}19 \text{ sau } 119\%;$$

$$\vdots$$

$$I_{10/9} = \frac{y_{10}}{y_9} = \frac{53\,721}{49\,063} = 1{,}095 = 109{,}5\%.$$

Şi între aceşti indicatori relativi există anumite relaţii care permit trecerea de la o formă la alta. Astfel, dacă se face produsul indicilor de creştere cu bază în lanţ se obţine indicele de creştere cu bază fixă al întregii perioade :

$$\prod_{i=0}^{n} I_{i/i-1} = I_{n/0}, \tag{10.7}$$

deci :

$$I_{1/0} \cdot I_{2/1} \cdot I_{3/2} \ldots I_{n/n-1} = I_{n/0};$$

$$1{,}208 \cdot 1{,}19 \cdot 1{,}124 \cdot 1{,}132 \cdot 1{,}114 \cdot 1{,}124 \cdot 1{,}101 \cdot 1{,}077 \cdot$$

$$\cdot 1{,}049 \cdot 1{,}095 = 3{,}126.$$

În mod corespunzător se poate face trecerea şi de la indicii de dinamică cu bază fixă la cei cu bază în lanţ. Dacă un indice de dinamică cu bază fixă al perioadei curente se raportează la indicele cu bază fixă al perioadei precedente se obţine indicele cu bază în lanţ corespunzător :

$$\frac{I_{i/0}}{I_{i-1/0}} = \frac{y_i}{y_0} : \frac{y_{i-1}}{y_0} = \frac{y_i}{y_{i-1}}.$$

Deci :

$$\frac{I_{i/0}}{I_{i-1/0}} = 1_{i/i-1} \tag{10.8}$$

De exemplu :

$$\frac{I_{2/0}}{I_{1/0}} = \frac{1{,}439}{1{,}208} = 1{,}19; \quad 1{,}19 = I_{2/1}.$$

Dar, în statistică interesează nu numai de cîte ori a crescut fenomenul cercetat în timp, ci şi cu cît nivelul comparat a depăşit în mărimi relative nivelul folosit ca bază de raportare (fixă sau variabilă). În acest caz se calculează ritmul de creştere, care poate fi şi el cu bază fixă sau cu bază în lanţ (variabilă).

Scanned with OKEN Scanner

Ritmul de creştere denumit şi ritmul sporului cu bază fixă se calculează ca raport între sporul cu bază fixă al fiecărei perioade şi nivelul anului de bază. Acest indicator se notează cu $R_{i/0}$ şi este înscris în col. 6 a tabelului 10.2. De regulă, şi acesta se exprimă sub formă de procente :

$$R_{i/0} = \frac{y_i - y_0}{y_0} \cdot 100 \qquad (10.9)$$

Deci, în exemplul precedent va fi :

$$R_{1/0} = \frac{y_1 - y_0}{y_0} \cdot 100 = \frac{3\,591}{17\,215} \cdot 100 = 20{,}8\%;$$

$$R_{2/0} = \frac{y_2 - y_0}{y_0} \cdot 100 = \frac{7\,554}{17\,215} \cdot 100 = 43{,}9\%;$$

$$\vdots$$

$$R_{10/0} = \frac{y_{10} - y_0}{y_0} \cdot 100 = \frac{36\,506}{17\,215} \cdot 100 = 2{,}12 \text{ sau } 212\%.$$

După cum se poate observa, indicii ritmului sporului cu bază fixă arată cu cît s-a depăşit relativ nivelul anului de bază ; cu alte cuvinte, se poate spune că acest indicator reprezintă expresia relativă a sporului realizat în fiecare an în raport cu baza întregii perioade, care s-a considerat ca fiind egală cu 100%. Acest lucru se poate demonstra şi mai sugestiv, dacă indicele de dinamică cu bază fixă este scris sub forma diferenţei a două rapoarte avînd acelaşi numitor, adică :

$$R_{i/0} = \frac{y_i - y_0}{y_0} = \frac{y_i}{y_0} - \frac{y_0}{y_0} = \frac{y_i}{y_0} - 1.$$

Ori, $\frac{y_i}{y_0}$ este egal tocmai cu indicele de dinamică cu baza fixă, deci, relaţia se mai poate scrie :

$$R_{i/0} = I_{i/0} - 1. \qquad (10.10)$$

Rezultă că dacă se cunoaşte indicele de dinamică cu bază fixă, indicele ritmului sporului este uşor de obţinut, şi anume, din indicele de dinamică cu bază fixă se scade 1 sau 100%, dacă primul este exprimat în procente, adică :

$$R_{1/0} = (I_{1/0} - 1)100 = (1{,}208 - 1)100 = 20{,}8\%;$$

$$R_{2/0} = (I_{2/0} - 1)100 = (1{,}439 - 1)100 = 43{,}9\%;$$

$$\vdots$$

$$R_{10/0} = (I_{10/0} - 1)100 = (3{,}12 - 1)100 = 212\%.$$

Scanned with OKEN Scanner

Ritmul de creştere cu bază în lanţ ($R_{i/i-1}$) se calculează ca raport între sporul cu bază în lanţ al fiecărui an şi nivelul anului precedent (col. 7, tabelul 10.2) şi se exprimă, de regulă, tot sub formă de procente :

$$R_{i/i-1} = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100. \qquad (10.11)$$

De exemplu :

$$R_{1/0} = \frac{y_1 - y_0}{y_0} \cdot 100 = \frac{3\,591}{17\,215} \cdot 100 = 20{,}8\%;$$

$$R_{2/1} = \frac{y_2 - y_1}{y_1} \cdot 100 = \frac{3\,963}{20\,806} \cdot 100 = 19\%;$$

$$\vdots$$

$$R_{10/9} = \frac{y_{10} - y_9}{y_9} \cdot 100 = \frac{4\,658}{49\,063} \cdot 100 = 9{,}5\%.$$

Aceeaşi relaţie se poate obţine şi între indicele de creştere şi indicele ritmului sporului cu bază în lanţ, respectiv :

$$R_{i/i-1} = \frac{y_i - y_{i-1}}{y_{i-1}} = \frac{y_i}{y_{i-1}} = \frac{y_{i-1}}{y_{i-1}} = \frac{y_i}{y_{i-1}} - 1,$$

deci :

$$R_{i/i-1} = I_{i/i-1} - 1. \qquad (10.12)$$

Dacă este exprimat în procente, atunci relaţia este :

$$R_{i/i-1} = (I_{i/i-1} \cdot 100) - 100.$$

De exemplu :

$$R_{1/0} = (I_{1/0} - 1) \cdot 100 = (1{,}208 - 1) \cdot 100 = 20{,}8\%;$$

$$R_{2/1} = (I_{2/1} - 1) \cdot 100 = (1{,}19 - 1) \cdot 100 = 19{,}0\%;$$

$$\vdots$$

$$R_{10/9} = (I_{10/9} - 1) \cdot 100 = (1{,}095 - 1) \cdot 100 = 9{,}5\%.$$

De subliniat este faptul că trecerea de la ritmul sporului cu bază fixă la cel cu bază în lanţ se poate face numai prin transformarea acestora în indici de dinamică corespunzători :

$$I_{i/0} = (R_{i/0} + 1) \cdot 100 \text{ şi}$$

$$I_{i/i-1} = (R_{i/i-1} + 1) \cdot 100, \qquad (10.13)$$

Scanned with OKEN Scanner

deoarece produsul indicilor ritmului sporului cu bază în lanţ nu este egal cu indicele ritmului cu bază fixă al întregii perioade :

$$\prod_{i=1}^{n} R_{i/i-1} \neq R_{n/0}. \qquad (10.14)$$

Folosind împreună indicatorii absoluţi şi relativi, se pot trage cîteva concluzii în legătură cu evoluţia producţiei de energie electrică a României între anii 1965—1975. Din indicatorii obţinuţi (vezi tabelul 10.2) se poate constata că producţia a crescut an de an, creşteri mai mari înregistrîndu-se, în special, în ultimii ani, începînd cu anul 1970, cînd producţia de energie electrică s-a dublat în comparaţie cu aceea a anului 1965, ajungînd în anul 1975 să fie de 3,12 ori mai mare. Dacă se urmăreşte an de an evoluţia producţiei, se poate constata că, în ultimii trei ani ai seriei, ritmul de creştere anuală a producţiei de energie electrică a scăzut.

Din compararea directă s-ar părea că cea mai mare creştere a producţiei s-ar fi realizat în 1966/1965 ($R = 20{,}8\%$) iar cea mai scăzută creştere în 1974/1973 ($R = 4{,}9\%$). Dar această comparaţie nu este concludentă, deoarece nivelul absolut al producţiei a crescut permanent. Rezultă că pentru interpretarea ritmurilor anuale de creştere trebuie avut în vedere următorul aspect : spre deosebire de indicii cu bază fixă, care sînt comparabili între ei avînd acelaşi numitor, indicii cu baza în lanţ avînd bază variabilă, nu se pot compara nemijlocit. Pentru efectuarea comparaţiei indicilor cu bază în lanţ, în scopul desprinderii tendinţei de dezvoltare a fenomenului studiat, este necesar să se calculeze un indicator intermediar care este valoarea absolută a unui procent de creştere cu bază fixă şi cu bază în lanţ.

Astfel, pentru a sublinia linia mereu ascendentă a economiei noastre naţionale este suficient să se arate că, pentru orice indice al venitului naţional calculat cu baza 1970, la un procent de creştere revin 2,121 miliarde lei, în timp ce pentru un indice cu baza 1975, în medie cîte 3,619 miliarde lei [1]. Deci, în 1975 faţă de 1970, valoarea absolută a unui procent de sporire a venitului naţional a crescut cu 71% $\left(\frac{3{,}619}{2{,}121}\right) \cdot 100\% - 100)$.

Valoarea absolută a unui procent de creştere cu bază fixă este aceeaşi pentru întreaga perioadă, deoarece nivelul care s-a considerat egal cu 100% este nivelul anului de bază (y_0) şi exprimă cîte unităţi, din sporul înregistrat într-un an, revin la fiecare procent din ritmul sporului.

Dacă se notează cu $A_{i/0}$ valoarea absolută a unui procent de creştere cu bază fixă, aceasta se va calcula ca raport între sporul cu bază fixă şi indicele ritmului sporului corespunzător aceleiaşi perioade, astfel :

$$A_{i/0} = \frac{\Delta_{i/0}}{(R_{i/0})\,100} \qquad (10.15)$$

---

[1] Anuarul Statistic al R.S. România, 1977, D.C.S., p. 84

Scanned with OKEN Scanner

$$A_{1/0} = \frac{\Delta_{1/0}}{(R_{1/0})100} = \frac{3\,591}{20,8} = 172,6 \text{ mil. Kwh};$$

$$A_{2/0} = \frac{\Delta_{2/0}}{(R_{2/0})100} = \frac{7\,554}{43,9} = 172,1 \text{ mil. Kwh};$$

$$\vdots$$

$$A_{10/0} = \frac{\Delta_{10/0}}{(R_{10/0})100} = \frac{36\,506}{2,12} = 172,2 \text{ mil. Kwh.}$$

Se observă că, indiferent pentru ce perioadă se calculează, valoarea absolută a unui procent de creştere cu bază fixă este aceeaşi, deoarece baza rămîne aceeaşi pentru toată perioada. (Diferenţele nesemnificative apar din cauza rotunjirilor de calcul.) Datorită faptului că perioada de bază (y_0) s-a considerat de fiecare dată egală cu 100%, mai simplu de aflat valoarea absolută ce revine unui procent de creştere cu bază fixă este să se facă raportul între nivelul anului de bază şi 100:

$$A_{n/0} = \frac{y_0}{100} = \frac{17\,215}{100} = 172,15 \text{ mil. Kwh.} \qquad (10.16)$$

Deci, aceşti indici sînt comparabili între ei şi se pot face cu ei operaţii de adunare şi scădere procentuală.

Valoarea absolută a unui procent de creştere cu bază în lanţ se va nota în mod corespunzător cu $A_{i/i-1}$ şi se bazează pe acelaşi raţionament, adică:

$$A_{i/i-1} = \frac{\Delta_{i/i-1}}{(R_{i/i-1})100};$$

$$A_{1/0} = \frac{\Delta_{1/0}}{(R_{1/0})\,100} = \frac{3\,591}{20,8} = 172,6; \qquad (10.17)$$

$$A_{2/1} = \frac{\Delta_{2/1}}{(R_{2/1})\,100} = \frac{3\,963}{19,0} = 208,6;$$

$$A_{3/2} = \frac{\Delta_{3/2}}{(R_{3/2})\,100} = \frac{3\,059}{12,4} = 246,7;$$

$$A_{4/3} = \frac{\Delta_{4/3}}{(R_{4/3})\,100} = \frac{3\,681}{13,2} = 278,9;$$

$$A_{5/4} = \frac{\Delta_{5/4}}{R_{5/4}} = \frac{3\,579}{11,4} = 313,9;$$

Scanned with OKEN Scanner

$$A_{6/5} = \frac{\Delta_{6/5}}{R_{6/5}} = \frac{4\,366}{12,4} = 352,1;$$

$$A_{7/6} = \frac{\Delta_{7/6}}{R_{7/6}} = \frac{3\,985}{10,1} = 394,6;$$

$$A_{8/7} = \frac{\Delta_{8/7}}{R_{8/7}} = \frac{3\,340}{7,7} = 433,8;$$

$$A_{9/8} = \frac{\Delta_{9/8}}{R_{9/8}} = \frac{2\,284}{4,9} = 466,1;$$

$$A_{10/9} = \frac{\Delta_{10/9}}{R_{10/9}} = \frac{4\,658}{9,5} = 490,3.$$

Se observă cu uşurinţă că valoarea absolută a unui procent de creştere cu baza în lanţ a sporit de la an la an — inclusiv în anul 1974, cînd s-a realizat ritmul cel mai scăzut — ca urmare a creşterii bazei de raportare de la un an la altul. Cum mărimea absolută a unui procent de creştere este proporţională cu baza de raportare, înseamnă că el se mai poate calcula ca a suta parte din nivelul anului precedent:

$$A_{i/i-1} = \frac{y_{i-1}}{100};$$

$$A_{1/0} = \frac{y_0}{100} = \frac{17\,215}{100} = 172,2;$$

$$A_{2/1} = \frac{y_1}{100} = \frac{20\,806}{100} = 208,1;$$

$$\vdots$$

$$A_{10/9} = \frac{y_9}{100} = \frac{49\,063}{100} = 490,6.$$

Se poate considera că acest indicator face legătura dintre indicatorii absoluţi şi cei relativi, ajutînd la interpretarea corectă a acestora.

10.3.3. Indicatorii medii ai unei serii cronologice de intervale

Prin calcularea indicatorilor absoluţi şi relativi s-au caracterizat relaţiile care există între termenii individuali ai unei serii cronologice, luaţi cîte doi. Aceşti indicatori arată tocmai gradul de variabilitate a ter-

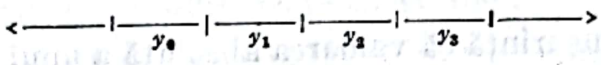
Scanned with OKEN Scanner

menilor unei serii cronologice, ca urmare a influenţei exercitate de toate cauzele şi condiţiile ce determină evoluţia fenomenului respectiv. Or, în caracterizarea evoluţiei fenomenelor trebuie să se determine şi tendinţele de dezvoltare a întregii serii, care apar ca rezultat al influenţei cauzelor esenţiale, prin anihilarea influenţelor factorilor cu caracter întîmplător, prezente în mărimea tuturor indicatorilor seriei. Aceasta, după cum se cunoaşte, se realizează cu ajutorul indicatorilor medii.

În analiza dinamicii fenomenelor se pot calcula medii de nivel şi medii de ritm. Mediile de nivel vor caracteriza nivelul mediu al termenilor unei serii cronologice şi nivelul mediu al sporului, iar mediile de ritm vor măsura indicele mediu al dinamicii şi indicele ritmului mediu de creştere.

Nivelul mediu al unei serii cronologice. La calculul nivelului mediu al unei serii dinamice trebuie ţinut seama de faptul că ele pot să apară sub formă de serii cronologice de intervale şi serii cronologice de momente.

Dacă se reprezintă pe o axă termenii unei serii cronologice de intervale, ei vor apărea astfel :



Fiecare termen se plasează în interiorul întregului interval. Termenii seriei, referindu-se la fiecare interval fără întrerupere, se pot cumula şi se obţine nivelul totalizat al caracteristicii pe întreaga perioadă $\left(\sum_{i=1}^{n} y_i\right)$.

Aceasta înseamnă că *nivelul mediu al unei serii dinamice de intervale* se poate calcula aplicînd formula mediei aritmetice simple (intervalele de timp fiind egale între ele), adică :

$$\bar{y} = \frac{\sum_{i=0}^{n} y_i}{n+1}. \qquad (10.19)$$

În exemplul prezentat $\Sigma y_i = 389\,671$ (vezi tabelul 10.2, totalul coloanei 1), iar $n + 1 = 11$, deci, nivelul mediu va fi :

$$\bar{y}_{1965-1975} = \frac{389\,671}{11} = 35\,424{,}64 \text{ mil. Kwh/anual.}$$

Deci, în perioada 1965—1975, producţia de energie electrică medie anuală, în ţara noastră reprezenta 35 424,64 mil. Kwh, cu o variaţie între 17 215 şi 53 721 mil. Kwh.

În mod analog se va calcula şi *sporul mediu anual* din sporurile anuale cu baza în lanţ, după formula :

$$\bar{\Delta} = \frac{\sum_{i=1}^{n} \Delta_{i/i-1}}{n} \qquad (10.20)$$

Scanned with OKEN Scanner

în care :

n reprezintă numărul sporurilor cu baza în lanţ ai seriei dinamice.

Într-o serie dinamică, numărul sporurilor cu baza în lanţ este mai mic cu o unitate faţă de numărul termenilor seriei. Deci, sporul mediu pentru exemplul luat :

$$\bar{\Delta}_{1965-1975} =$$

$$= \frac{3\,591+3\,963+3\,059+3\,681+3\,579+4\,366+3\,985+3\,340+2\,284+4\,658}{10} =$$

$$= \frac{36\,506}{10} = 3\,650{,}6 \text{ mil. Kwh/anual.}$$

Aceasta înseamnă că, în medie, dacă producţia creştea sub formă liniară, pe seama unor cauze cu influenţă constantă pe toată perioada (1965—1975), atunci, an de an, producţia de energie electrică a R.S. România trebuia să crească cu 3 650 mil. Kwh, calculată faţă de nivelul anului precedent.

Dar, după cum se observă, sporul mediu anual depinde, în ultimă instanţă, de diferenţa de mărime dintre primul şi ultimul termen, deoarece, pe baza relaţiei ce există între sporurile cu baza în lanţ şi cele cu bază fixă, suma sporurilor cu baza în lanţ este egală cu sporul cu baza fixă al întregii perioade :

$$\bar{\Delta} = \frac{\sum_{i=1}^{n} \Delta_{i/i-1}}{n} = \frac{\Delta_{n/0}}{n},$$

de unde :

$$\bar{\Delta} = \frac{y_n - y_0}{n}, \qquad (10.21)$$

adică :

$$\bar{\Delta}_{1961-1971} = \frac{53\,721 - 17\,215}{10} = \frac{36\,506}{10} =$$

$$= 3\,650{,}6 \text{ mil. Kwh/anual.}$$

Rezultă că, în acest caz, se pot folosi două formule de calcul pentru sporul mediu (10.20 şi 10.21).

Se poate trage concluzia că sporul mediu are sens economic numai în măsura în care între sporurile cu baza în lanţ nu există variaţie mare şi prezintă aceeaşi tendinţă pe toată perioada ; altfel, prin compensarea abaterilor în plus şi minus din interiorul seriei, se va estompa neomogenitatea datelor prezentate. Dacă în interiorul aceleiaşi serii se întîlnesc tendinţe opuse, care, pe grafic, corespund unei schimbări de forma unei parabole de gradul doi, cu un punct de maxim sau de minim, atunci seria

Scanned with OKEN Scanner

trebuie să se dividă în două părţi, calculînd separat indicatorii medii respectivi.

Indicele mediu de creştere trebuie să arate de cîte ori trebuie să crească an de an nivelul fenomenului cercetat, dacă el ar fi crescut de forma unei progresii geometrice a cărei raţie să exprime influenţa constantă a factorilor esenţiali pe întreaga perioadă. Aceasta ar corespunde unor creşteri absolute din ce în ce mai mari, ca urmare a mărimii nivelului de dezvoltare a fenomenului cercetat, de la o perioadă la alta, iar indicii individuali de creştere cu baza în lanţ ar fi fost egali între ei. Indicii cu baza în lanţ, fiind egali între ei, ar fi fost egali cu indicele mediu, iar produsul lor ar fi fost egal cu indicele de creştere cu bază fixă :

$$I_{1/0} \cdot I_{2/1} \cdot I_{3/2} \cdot I_{4/3} \ldots I_{n/n-1} = I_{n/0}; \qquad (10.22)$$

$$\bar{I} \cdot \bar{I} \cdot \bar{I} \ldots \bar{I} = I_{n/0}$$

de unde :

$$I_{1/0} \cdot I_{2/1} \cdot I_{3/2} \cdot I_{4/3} \ldots I_{n/n-1} = \bar{I} \cdot \bar{I} \cdot \bar{I} \ldots \bar{I}.$$

$\bar{I}$ fiind o mărime constantă, produsul din membrul doi al egalităţii se poate înlocui cu valoarea sa ridicată la o putere egală cu numărul de termeni ($\bar{I}^n$).

$$I_{1/0} \cdot I_{2/1} \cdot I_{3/2} \cdot I_{4/3} \ldots I_{n/n-1} = \bar{I}^n$$

sau mai restrîns :

$$\prod_{i=1}^{n} I_{i/i-1} = \bar{I}^n, \qquad (10.23)$$

de unde :

$$\bar{I} = \sqrt[n]{\prod_{i=1}^{n} I_{i/i-1}}. \qquad (10.24)$$

Şi în acest caz numărul indicilor cu baza în lanţ este mai mic cu o unitate faţă de termenii seriei iniţiale. În acelaşi timp se poate observa că formula folosită corespunde, de fapt, formulei de calcul a mediei geometrice (6.39), în care :

$\bar{x}_g$ reprezintă indicele mediu ;

x_i — indicii din care se calculează media.

Aplicată la exemplul anterior, va fi :[1]

$$\bar{I}_{1965-1975} =$$

$$= \sqrt[10]{1{,}208 \cdot 1{,}19 \cdot 1{,}124 \cdot 1{,}132 \cdot 1{,}114 \cdot 1{,}124 \cdot 1{,}101 \cdot 1{,}077 \cdot 1{,}049 \cdot 1{,}095}.$$

[1] De subliniat că indicii de creştere trebuie luaţi sub formă de coeficienţi şi nu sub formă de procente, altfel fiecare termen ar apare de două ori şi, deci, n-ar mai corespunde indicelui radicalului.

Scanned with OKEN Scanner

Pentru a rezolva acest calcul este necesar să se aplice logaritmii. Prin aplicarea logaritmilor, media geometrică se transformă într-o medie aritmetică a logaritmilor factorilor, adică:

$$\log \bar{I} = \frac{0{,}08207 + 0{,}07555 + 0{,}05077 + 0{,}05385 + 0{,}04689 +}{10}$$

$$\frac{+ 0{,}05077 + 0{,}04179 + 0{,}03222 + 0{,}02078 + 0{,}03941}{10} = 0{,}04941,$$

de unde, antilogaritmînd:

$$\bar{I}_{1965-1975} = 1{,}1205 \simeq 112{,}1\%.$$

Rezultă că, în medie, an de an, dacă nu influenţau şi factorii întîmplători, producţia de energie electrică pe perioada 1965—1975 ar fi crescut de forma unei progresii geometrice, cu raţia de 1,121, în aşa fel încît pornind de la 17 215 mil. Kwh în 1965 să se ajungă la o producţie de 53 721 mil. Kwh în 1975.

Ca şi în cazul sporului mediu, indicele mediu de creştere calculat prin media geometrică ascunde variaţiile din interiorul seriei de la unitate de timp la alta, ajungînd astfel ca mărimea sa să depindă, de fapt, numai de raportul dintre termenii extremi şi lungimea seriei. Din relaţia arătată mai înainte, produsul indicilor cu bază în lanţ este egal cu indicele cu bază fixă al întregii perioade:

$$\bar{I} = \sqrt[n]{\prod_{i=1}^{n} I_{i/i-1}} = \sqrt[n]{I_{i/0}} = \sqrt[n]{\frac{y_n}{y_0}} \qquad (10.25)$$

Ca şi în cazul sporului mediu, şi pentru indicele mediu de creştere se pot folosi două formule de calcul (10.24 şi 10.25).

Ele se folosesc în funcţie de datele de care se dispune. Pentru seriile mai lungi este mai avantajos de calculat prin cea de-a doua formulă.

Aplicînd şi cea de-a doua formulă la exemplul luat, se va obţine:

$$\bar{I}_{1965-1975} = \sqrt[10]{\frac{53\,721}{17\,215}}.$$

Efectuînd calculele necesare se obţine aceeaşi valoare a indicelui mediu de dinamică ($\bar{I}_{1965-1975} \approx 1{,}121$).

În legătură cu calculul indicelui mediu, în practică se poate pune şi problema ca, dispunînd de mai mulţi indici medii ce caracterizează mai multe subperioade succesive de timp, să se calculeze indicele general ce caracteri-

Scanned with OKEN Scanner

zează întreaga perioadă. Acesta se poate calcula ca medie geometrică ponderată a indicilor medii de creştere :

$$\bar{\bar{I}} = \sqrt[\sum_{i=1}^{k} n_i]{\bar{I}_1^{n_1} \cdot \bar{I}_2^{n_2} \cdot \bar{I}_3^{n_3} \dots \bar{I}_k^{n_k}} \qquad (10.26)$$

în care :

$\bar{\bar{I}}$ reprezintă indicele mediu general de creştere ;
$\bar{I}$ — indicele mediu parţial de creştere ;
n_i — numărul indicilor cu bază în lanţ ce intră în componenţa fiecărui indice mediu parţial ;
k — numărul subperioadelor, adică al indicilor medii parţiali.

Ritmul mediu de creştere arată cu cît a crescut fenomenul respectiv în mărimi relative, pe perioada analizată, în medie de la o unitate de interval la alta.

Ritmul mediu de creştere se calculează pe baza relaţiei ce există între indicele de dinamică şi ritmul de creştere care se păstrează şi în cadrul indicatorilor medii. Deci, *ritmul mediu de creştere* se calculează ca diferenţă între indicele mediu de dinamică, exprimat în procente şi 100% (care reprezintă baza de raportare).

$$\bar{R} = (\bar{I} \times 100) - 100 ; \qquad (10.27)$$

$$\bar{R} = (1{,}121 \times 100) - 100 = 12{,}1\%.$$

Deci, în medie, producţia de energie electrică a R.S. România a crescut, între anii 1965—1975, cu un ritm mediu anual de 12,1%, ceea ce reprezintă unul din ritmurile înalte nu numai în ţara noastră, ci şi pe plan mondial.

În concluzie, indicatorii obţinuţi prin prelucrarea unei serii cronologice pot fi constituiţi într-un sistem, în care fiecare indicator permite să se scoată în evidenţă cîte un aspect al modului de dezvoltare a fenomenelor cercetate. Veridicitatea acestor indicatori este determinată de modul de alcătuire a seriilor cronologice, de omogenitatea datelor empirice utilizate. În cazul seriilor cronologice neomogene, cu tendinţe variate de dezvoltare, se impune calcularea acestor indicatori pe fiecare etapă sub formă de indicatori parţiali, altfel conţinutul indicatorilor obţinuţi din prelucrare nu este real şi concluziile teoretice şi practice formulate n-au un fundament corect şi nu pot fi folosite pentru calculele de prognoză.

10.4. PRELUCRAREA SERIILOR CRONOLOGICE DE MOMENTE

În cazul unei serii de momente se pot întîlni două situaţii : serii de momente cu intervale egale între momente, şi serii de momente cu intervale neegale între ele.

Scanned with OKEN Scanner

Dacă intervalele dintre momentul înregistrării termenilor sînt egale între ele, atunci prelucrarea seriei se poate face în două variante :

1) Se transformă seria de momente în serie de intervale calculîndu-se media aritmetică simplă pentru fiecare interval şi apoi se calculează indicatorii absoluţi, relativi şi medii prezentaţi anterior. De exemplu, avînd valorile fondurilor fixe înregistrate la sfîrşitul fiecărui an, se calculează valoarea medie anuală a acestora şi se obţine o nouă serie care, faţă de prima serie, va avea cu un termen mai puţin. Datele din noua serie, fiind comparabile din punctul de vedere al timpului la care se referă, pot fi comparate între ele atît în mărime absolută, cît şi în mărime relativă. Rezultatele obţinute din prelucrarea unor astfel de serii se pot include în calcule de corelaţie cu seriile unor caracteristici cu care se găsesc în interdependenţă, de exemplu, în analiza corelaţiei realizată între dinamica fondurilor fixe şi dinamica producţiei.

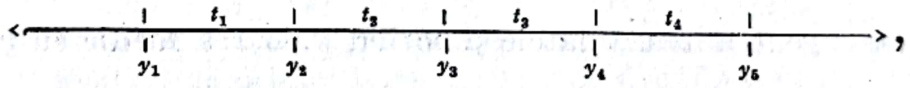
2) Seriile de momente cu intervale egale între termenii înregistraţi se prelucrează ca atare şi atunci se pot obţine indicatorii absoluţi, relativi şi medii prezentaţi anterior, cu excepţia nivelului mediu care se calculează folosind o formă specială de medie aritmetică, cunoscută în literatura de specialitate ca medie cronologică.

10.4.1. Media cronologică simplă

Media cronologică simplă se aplică pentru seriile de momente cu intervale egale între ele.

Dacă intervalele între momentele înregistrării termenilor sînt egale atunci prelucrarea seriei se face ca şi în cazul unei serii de intervale de timp, cu excepţia nivelului mediu, care se calculează cu ajutorul mediei cronologice simple.

Reprezentarea termenilor pe o axă a unei serii de momente cu intervale egale între ele, va fi :



în care :

y reprezintă termenii seriei dinamice care iau valori de la y_1 la y_n;

t — intervalele între momentele seriei care pot lua valori de la t_1 la t_{n-1}.

Din axa valorilor y_i rezultă că termenii sînt echidistanţi între ei şi că fiecare marchează cîte un moment al perioadei analizate. Cum într-o serie de momente sînt n termeni şi $n-1$ intervale, înseamnă că fiecare interval va fi marcat de cîte doi termeni. Din această cauză termenii extremi apar o singură dată, iar termenii intermediari de cîte două ori. De exemplu, y_2 marchează şi sfîrşitul primului interval şi începutul celui de-al doilea, pe cînd y_1 apare o singură dată marcînd începutul primului interval.

Deci, în acest caz, pentru a putea calcula media pe total este necesar să se calculeze mediile parţiale pe fiecare interval, care se vor calcula ca

Scanned with OKEN Scanner

medii aritmetice simple din cei doi termeni ce marchează intervalul respectiv.

Pentru a ilustra modul de calcul se ia un exemplu, cu date convenționale, cu privire la stocurile de mărfuri într-o unitate comercială, în cursul trimestrului I al unui an oarecare :

1 ianuarie 500 000 lei (y_1) ;

1 februarie 580 000 lei (y_2) ;

1 martie 620 000 lei (y_3) ;

1 aprilie 640 000 lei (y_4).

Pentru a calcula media trimestrială se vor calcula, mai întîi, mediile pe fiecare lună în parte :

$$\text{ianuarie} = \frac{y_1 + y_2}{2} = \frac{500\,000 + 580\,000}{2} = 540\,000 \text{ lei};$$

$$\text{februarie} = \frac{y_2 + y_3}{2} = \frac{580\,000 + 620\,000}{2} = 600\,000 \text{ lei};$$

$$\text{martie} = \frac{y_3 + y_4}{2} = \frac{620\,000 + 640\,000}{2} = 630\,000 \text{ lei}.$$

Media pe trimestru se va calcula ca medie generală :

$$\bar{y} = \frac{540\,000 + 600\,000 + 630\,000}{3} = \frac{1\,770\,000}{3} = 590\,000 \text{ lei}.$$

Dacă se generalizează datele și notînd valoarea medie cu $\bar{y}_{cr}$ se va obține :

$$\bar{y}_{cr} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \ldots + \frac{y_{n-1} + y_n}{2}}{n - 1}.$$

Se iau $n - 1$ termeni la numitor, deoarece față de numărul termenilor seriei, se pot calcula $n - 1$ medii parțiale, pe fiecare interval.

Dacă în formula mediei ($\bar{y}$) se efectuează sumele de la numărător, se va obține :

$$\bar{y}_{cr} = \frac{\frac{y_1 + 2y_2 + 2y_3 + \ldots + y_n}{2}}{n - 1},$$

Scanned with OKEN Scanner

în continuare simplificînd cu 2 la fracţia de la numărător :

$$\bar{y}_{cr} = \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots + \frac{y_n}{2}}{n-1} \qquad (10.28)$$

Această formulă, aplicîndu-se numai la seriile cronologice de momente, se numeşte *medie cronologică*, avînd acelaşi conţinut ca media generală calculată din medii parţiale.

Deoarece intervalele dintre momente sînt egale, ea apare ca *medie cronologică simplă* şi se defineşte ca semisuma termenilor extremi plus suma termenilor intermediari raportate la numărul termenilor seriei minus unul, aceasta şi datorită faptului că termenii extremi se iau pe jumătate şi, deci, împreună fac un singur termen.

Aplicată direct, fără a calcula în prealabil mediile pe fiecare interval, va fi :

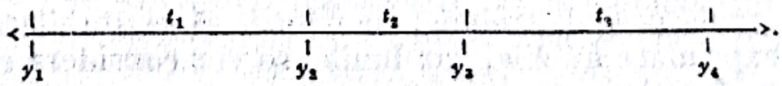
$$\bar{y}_{cr} = \frac{\frac{500\,000}{2} + 580\,000 + 620\,000 + \frac{640\,000}{2}}{3} = \frac{1\,770\,000}{3},$$

$$\bar{y}_{cr} = 590\,000 \text{ lei.}$$

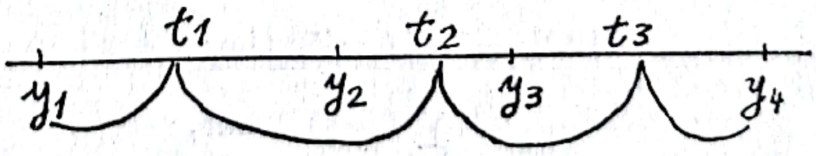
Deci, s-a ajuns la aceeaşi valoare numerică ca şi în cazul cînd această medie s-a calculat ca o medie generală din cele trei medii parţiale lunare.

10.4.2. Media cronologică ponderată

Se foloseşte în cazul în care între momentele seriei există intervale neegale. Atunci, pe o axă, termenii se vor distribui astfel :



Termenii fiind distribuiţi la intervale neegale, pentru a calcula media se va considera că modificarea lor se realizează uniform de la un interval la altul. În acest caz fiecare termen va fi proporţional cu cîte o jumătate din mărimea intervalelor alăturate, astfel :



Scanned with OKEN Scanner

Deci, pe baza acestui raţionament se vor considera ca ponderi pentru fiecare termen cîte o jumătate din mărimea intervalelor alăturate, folosindu-se astfel o *medie cronologică ponderată* :

$$\bar{y}_{cr} = \frac{y_1\left(\frac{t_1}{2}\right) + y_2\left(\frac{t_1}{2} + \frac{t_2}{2}\right) + \ldots + y_n\left(\frac{t_{n-1}}{2}\right)}{\left(\frac{t_1}{2}\right) + \left(\frac{t_1}{2} + \frac{t_2}{2}\right) + \ldots + \left(\frac{t_{n-1}}{2}\right)}$$

$$\bar{y}_{cr} = \frac{y_1\left(\frac{t_1}{2}\right) + y_2\left(\frac{t_1 + t_2}{2}\right) + \ldots + y_n\left(\frac{t_{n-1}}{2}\right)}{\frac{t_1}{2} + \frac{t_1 + t_2}{2} + \ldots + \frac{t_{n-1}}{2}}$$

$$\bar{y}_{cr} = \frac{y_1\left(\frac{t_1}{2}\right) + y_2\left(\frac{t_1 + t_2}{2}\right) + \ldots + y_n\left(\frac{t_{n-1}}{2}\right)}{\sum_{i=1}^{n-1} t_i}. \qquad (10.29)$$

Aici primul şi ultimul termen se ponderează cu jumătate din primul interval de timp (respectiv ultimul interval), iar termenii intermediari cu cîte o jumătate din intervalele alăturate.

De exemplu, stocurile de mărfuri existente într-o unitate comercială (date convenţionale) într-un semestru au fost :

Data	Stocul existent (mii lei)
1 ianuarie	500 000
1 februarie	580 000
15 martie	680 000
10 mai	700 000
1 iulie	750 000

Fiind intervale neegale trebuie, mai întîi, să se stabilească mărimea intervalelor, exprimate în zile, iar lunile se vor considera ca fiind egale 30 zile fiecare, deci :

$$t_1 = 1.\text{I} - 1.\text{II} = 30 \text{ zile};$$

$$t_2 = 1.\text{II} - 15.\text{III} = 45 \text{ zile};$$

$$t_3 = 15.\text{III} - 10.\text{V} = 55 \text{ zile};$$

$$t_4 = 10.\text{V} - 1.\text{VII} = 50 \text{ zile};$$

$$\sum_{i=1}^{n-1} t_i = 180 \text{ zile},$$

Scanned with OKEN Scanner

de unde, nivelul mediu, calculat în mil. lei prin media cronologică ponderată, va fi:

$$\bar{y}_{cr} =$$

$$= \frac{500\left(\frac{30}{2}\right)+580\left(\frac{30+45}{2}\right)+680\left(\frac{45+55}{2}\right)+700\left(\frac{55+50}{2}\right)+750\left(\frac{50}{2}\right)}{\frac{30}{2}+\frac{30+45}{2}+\frac{45+55}{2}+\frac{55+50}{2}+\frac{50}{2}} =$$

$$= \frac{118\,750}{180} = 659{,}772 \text{ milioane lei.}$$

Deci:

$$\bar{y}_{cr} \simeq 660 \text{ mil. lei.}$$

De reţinut că, pentru seria de momente cu intervale neegale între datele înregistrate, media cronologică ponderată este singurul indicator ce caracterizează seria. Ceilalţi indicatori nu se pot calcula ca un sistem corelat, deoarece nu se îndeplineşte condiţia de uniformitate a periodicităţii cu care sînt prezentate datele seriei, adică nu sînt comparabile din punctul de vedere al variaţiei de timp. Aceste serii se folosesc, în special, pentru prezentarea datelor care provin din observări special organizate, de forma recensămintelor sau a inventarierii, şi prelucrarea lor se face în funcţie de specificul fiecărei serii.

10.5. AJUSTAREA SERIILOR CRONOLOGICE

Seriile cronologice, referindu-se la fenomenele social-economice, pot prezenta variaţii foarte mari de la o perioadă la alta. Aceste variaţii pot arăta tendinţe variate, dar care, de fapt, corespund legităţii interne, ce determină dezvoltarea fenomenelor respective.

Specific fenomenelor sociale este faptul că acţiunea sistematică a cauzelor esenţiale este modificată de influenţa cauzelor neesenţiale, nu numai în interiorul aceleiaşi perioade de timp, ci şi în dinamică. Fenomenele social-economice, fiind influenţate în dinamică, nu numai de acţiunea factorilor esenţiali, prezintă uneori abateri mai mari sau mai mici de la linia generală teoretică, care, în sens statistic, exprimă legitatea de dezvoltare specifică fiecărei etape.

La analiza unor serii dinamice se poate desprinde fie tendinţa de apropiere către aceeaşi valoare a creşterilor absolute înregistrate de la an la an, fie tendinţa de apropiere a ritmurilor de creştere cu baza în lanţ.

Dacă se reprezintă grafic datele seriei în funcţie de timp, atunci se poate observa că unele fenomene se pot dezvolta de formă rectilinie, altele de formă curbilinie.

Dacă seria de date se referă la o perioadă mai lungă de dezvoltare, se pot întîlni şi cazuri cînd o tendinţă de dezvoltare rectilinie se transformă într-o tendinţă de dezvoltare exponenţială, transformîndu-se ulterior într-o tendinţă rectilinie.

Scanned with OKEN Scanner

De asemenea, alte serii dinamice prezintă unele variaţii repetabile (cu caracter ciclic) în funcţie de modificarea anotimpurilor, iar mai rar se întîlnesc serii dinamice ai căror termeni par să se distribuie cu totul independent unul faţă de altul. Ori, în cadrul unei serii dinamice, termenii seriei nu pot fi cu totul independenţi între ei, ca în cazul seriilor de distribuţie, unde fiecare termen este independent faţă de celălalt şi este legat de o anumită probabilitate de apariţie, concretizată în frecvenţa sa de manifestare. După cum s-a mai arătat, în cadrul seriilor dinamice interdependenţa dintre termenii seriei trebuie privită în sensul că, aici, fiecare termen depinde, într-o anumită măsură, de nivelul precedent. De exemplu, producţia lunii februarie nu este cu totul independentă de condiţiile şi, respectiv, de nivelul producţiei din luna ianuarie a aceluiaşi an. Mai mult, analizînd cauzele şi condiţiile de bază ale producerii fenomenelor, se poate constata că ele se modifică sistematic, de la o unitate la alta, ca urmare a unor acumulări cantitative treptate ce pot fi urmărite în cadrul unei etape întregi de dezvoltare. Această interdependenţă obiectivă dintre termenii unei serii dinamice imprimă evoluţiei fenomenelor cercetate o anumită tendinţă, care, datorită faptului că fenomenele sînt influenţate din dinamica lor şi de factorii întîmplători, nu poate fi cunoscută decît printr-o analiză teoretică şi practică a tuturor termenilor seriei respective.

Statistica, prin metodele sale specifice, trebuie să studieze care este tendinţa de dezvoltare, cunoscută în literatura de specialitate şi sub denumirea de trend şi, prin ajustare, să se încerce separarea influenţei factorilor esenţiali, cu acţiune sistematică, de acţiunea factorilor accidentali, care fac ca între termenii empirici şi cei teoretici să existe abateri. Cu cît influenţa factorilor accidentali este mai pronunţată, cu atît gradul de variaţie de la o unitate de timp la alta este mai mare şi linia (curba) de tendinţă este mai dificil de identificat.

Această problemă se rezolvă, practic, prin operaţia de ajustare.

Prin ajustarea termenilor unei serii de date statistice, în sensul cel mai larg se înţelege operaţia de înlocuire a termenilor reali cu termeni teoretici ce exprimă legitatea specifică de dezvoltare obiectivă a fenomenelor la care se referă datele. În cazul seriilor dinamice legitatea de dezvoltare se analizează în funcţie de timp. Dar, variabila de timp nu trebuie considerată în sine ca factor determinant, ci ca un mijloc de sintetizare în mod succesiv a influenţei sistematice a factorilor care acţionează în cadrul aceloraşi condiţii de bază, dar cu dimensiuni diferite.

În cazul ajustării seriilor dinamice dispersia totală, care sintetizează mărimea medie a variaţiei produsă de influenţa tuturor factorilor, se descompune în dispersia calculată pe baza variaţiei termenilor reali de la valorile ajustate în funcţie de timp, plus dispersia calculată pe baza variaţiei acestor valori ajustate de la media termenilor seriei dinamice.

Dispersia totală (σ_y^2) se calculează după formula:

$$\sigma_y^2 = \frac{\Sigma (y_i - \bar{y})^2}{n}. \qquad (10.30)$$

Dispersia termenilor seriei de la valorile ajustate ($\sigma_{y/z}^2$) sintetizează influenţa factorilor reziduali — factori neînregistraţi — care, în cazul serii-

Scanned with OKEN Scanner

lor dinamice, sînt toți factorii în afară de factorul timp și se calculează cu formula :

$$\sigma^2_{y/z} = \frac{\Sigma(y_i - Y_{t_i})^2}{n}. \qquad (10.31)$$

în care :

Y_{t_i} reprezintă valoarea teoretică a variabilei y în funcție de timp.

Dispersia valorilor ajustate de la valoarea medie ($\sigma^2_{y/t}$), care sintetizează variația produsă numai de modificarea factorului timp :

$$\sigma^2_{y/t} = \frac{\Sigma(Y_{t_i} - \bar{y})^2}{n}. \qquad (10.32)$$

De subliniat este faptul că valorile teoretice (ajustate) în funcție de timp se pot stabili folosind mai multe procedee de calcul.

Condiția esențială a aplicării corecte a unui procedeu sau altuia de ajustare este ca numărul termenilor seriei să fie suficient de mare pentru a putea intra astfel în cîmpul de acțiune al legii numerelor mari, asigurînd astfel o compensare reală a abaterilor întîmplătoare. Este cunoscut, de asemenea, faptul că legitatea de dezvoltare a unui fenomen nu se poate urmări decît în cadrul unei etape întregi, în cadrul căreia condițiile de bază ale producerii fenomenelor se schimbă numai cantitativ. Aceasta înseamnă că numai atunci cînd această condiție — număr suficient de mare de termeni — este satisfăcută, se poate desprinde în mod veridic tendința (trendul) de dezvoltare a fenomenului respectiv.

În teoria și practica statistică se folosesc mai frecvent următoarele metode și procedee de ajustare :

— ajustarea pe baza mediilor mobile ;
— ajustarea prin metoda grafică ;
— ajustarea pe baza sporului mediu ;
— ajustarea pe baza indicelui mediu de creștere ;
— ajustarea prin metode analitice de calcul bazate pe procedeul celor mai mici pătrate.

10.5.1. Ajustarea pe baza mediilor mobile

Acest procedeu de ajustare se folosește, în special, acolo unde variația termenilor unei serii dinamice prezintă un aspect de regularitate ciclică. Prin calcularea mediilor mobile se înlătură această variație ciclică și se prezintă seria de date cu o variație lină, continuă.

Mediile mobile sînt medii parțiale, calculate dintr-un număr prestabilit de termeni, în care se înlocuiește pe rînd primul termen cu termenul ce urmează în seria care trebuie să fie ajustată.

Din această cauză, mediile mobile se mai numesc și medii glisante sau alunecătoare.

Scanned with OKEN Scanner

Se consideră, de exemplu, o serie formată din 8 termeni (y_i), care urmează să fie ajustaţi prin procedeul mediilor mobile ($\bar{y}_i$), calculate din cîte trei termeni.

Mediile mobile vor fi:

$$\bar{y}_1 = \frac{y_1 + y_2 + y_3}{3}; \qquad \bar{y}_4 = \frac{y_4 + y_5 + y_6}{3};$$

$$\bar{y}_2 = \frac{y_2 + y_3 + y_4}{3}; \qquad \bar{y}_5 = \frac{y_5 + y_6 + y_7}{3};$$

$$\bar{y}_3 = \frac{y_3 + y_4 + y_5}{3}; \qquad \bar{y}_6 = \frac{y_6 + y_7 + y_8}{3}.$$

Ajustarea, în acest caz, se va face după următorul model:

Valori empirice (y_i)	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
		→	↑	↑	↑	↑	↑	
Valori ajustate $\bar{y}_i$	—	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	$\bar{y}_5$	y_6	

Se observă că s-a obţinut un număr mai mic de medii mobile decît numărul termenilor seriei empirice, şi anume, din 8 termeni s-au obţinut 6 medii mobile. Aceasta înseamnă, în general, că se obţin atîtea medii mobile cîţi termeni are seria, mai puţin cu numărul termenilor din care s-au calculat mediile, micşorat cu o unitate. Dacă n este numărul termenilor seriei şi numărul termenilor pentru fiecare medie mobilă (n'), atunci numărul mediilor mobile va fi: $n-(n'-1)$.

În acest caz operaţia de ajustare a fost relativ uşoară, deoarece calculînd mediile dintr-un număr impar de termeni, fiecare medie s-a plasat în dreptul unui termen real al seriei şi care *a corespuns cu poziţia termenului central*. Dacă, însă, se calculează medii mobile dintr-un număr par de termeni, atunci fiecare medie mobilă se va plasa la mijlocul termenilor, deci, între cei doi termeni centrali. Pentru a putea face ajustarea termenilor în aceste condiţii se vor calcula mediile mobile din cîte doi termeni ai seriei ajustate. În această situaţie, mediile mobile iniţiale se vor numi medii mobile provizorii (notate cu $\bar{y}_i$), iar în a doua etapă se vor obţine medii mobile definitive sau centrate (notate cu $\bar{\bar{y}}_i$):

$$\bar{\bar{y}}_1 = \frac{\bar{y}_1 + \bar{y}_2}{2}; \qquad \bar{\bar{y}}_2 = \frac{\bar{y}_2 + \bar{y}_3}{2} \text{ etc.}$$

Ajustarea termenilor se va face după următorul model (seria din 8 termeni, medii mobile din cîte patru termeni):

Scanned with OKEN Scanner

Valori empirice (y_t)	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
medii mobile provizorii ($\bar{y}_t$)	—	—	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	$\bar{y}_5$	—	—
medii mobile definitive ($\bar{\bar{y}}_t$)	—	—	$\bar{\bar{y}}_1$	$\bar{\bar{y}}_2$	$\bar{\bar{y}}_3$	$\bar{\bar{y}}_4$		—	—

Deci, în cazul în care ajustarea se face pe baza mediilor mobile calculate din număr par de termeni, atunci mediile mobile se obţin în două trepte :

1) medii mobile provizorii ($\bar{y}_t$), care se plasează între termenii seriei;
2) medii mobile definitive sau centrate ($\bar{\bar{y}}_t$), care se plasează în dreptul termenilor seriei şi cu care se face ajustarea termenilor seriei iniţiale. Un exemplu numeric se va lua pentru calcularea gradului de sezonalitate (vezi paragraful 10.6.2).

10.5.2. Ajustarea pe baza metodei grafice

Se reprezintă grafic seria de date empirice şi apoi se trasează, vizual, dreapta sau curba care unește cele două puncte extreme ale seriei cronologice, astfel încît să aibă abateri minime faţă de poziţia valorilor reale în grafic. Această ajustare vizuală se bazează pe ipoteza că acţiunea tuturor cauzelor ar fi fost constantă pe toată perioada, imprimînd tuturor ter-

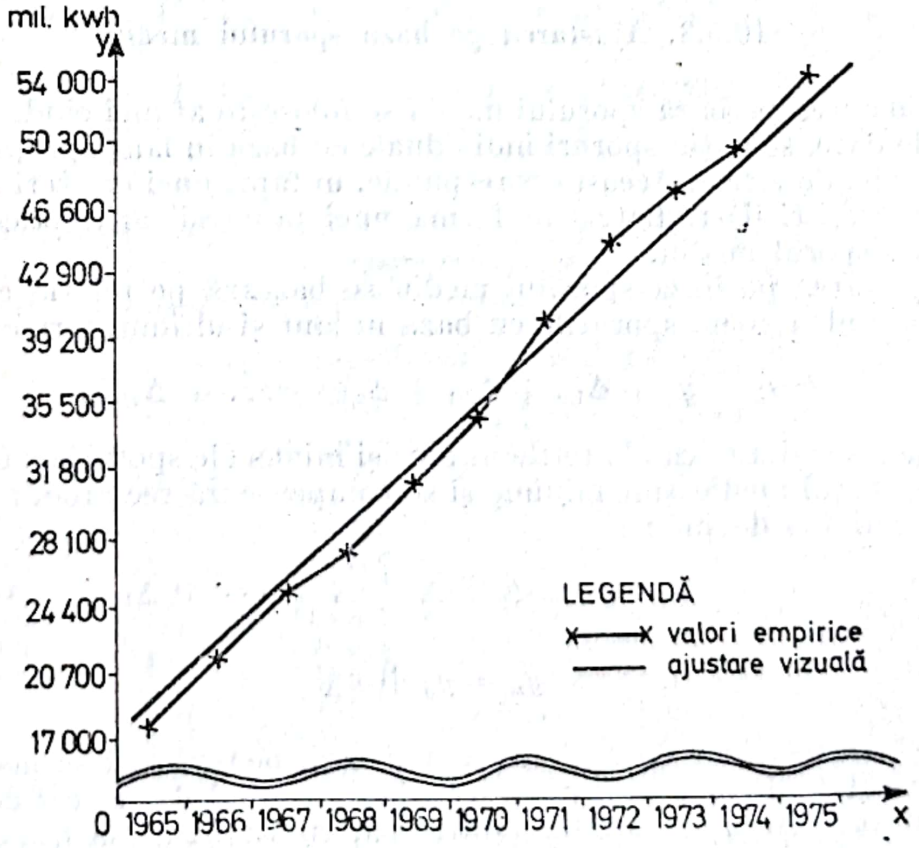


Fig. 10.1. Producţia de energie electrică în R.S. România în anii 1965—1975.

Scanned with OKEN Scanner

menilor aceeaşi formă de creştere absolută sau relativă şi care poate fi interpretată pe baza liniei (curbei) valorilor reale luate în funcţie de timp.

Graficul folosit pentru reprezentarea unei serii dinamice este cronograma[1], bazat pe sistemul coordonatelor rectangulare, în care timpul este trecut pe axa absciselor (OX), iar indicatorii urmăriţi sînt folosiţi pentru stabilirea scării de reprezentare pe axa ordonatelor (OY).

După cum se poate urmări pe grafic, producţia de energie electrică între 1965—1975, prezintă pentru cîţiva termeni tendinţă liniară, iar pentru ceilalţi tendinţă exponenţială. Aceasta sugerează necesitatea de a folosi mai multe procedee de ajustare şi alegerea aceluia care se apropie cel mai mult de tendinţa reală de evoluţie.

Rezultă că ajustarea pe baza reprezentării grafice constituie un instrument de apreciere a tendinţei de dezvoltare, în funcţie de care se poate alege însăşi metoda (procedeul) ce trebuie folosită în estimarea tendinţei de lungă şi scurtă durată.

În general, se acceptă ca cel mai bun mijloc de ajustare, acel procedeu care, aplicat la seria de date empirice, permite obţinerea unor termeni teoretici care dă abateri minime de la valorile reale corespunzătoare. În cazul de faţă, ţinînd seama de aliura graficului, este de aşteptat ca valorile teoretice care corespund ajustării de forma unei progresii aritmetice (sporul mediu şi ecuaţia liniei drepte) să dea cele mai mici abateri. Pentru verificare se vor folosi mai multe procedee de calcul şi se va alege cel care satisface cel mai bine condiţia de minim.

10.5.3. Ajustarea pe baza sporului mediu

Ajustarea pe baza sporului mediu se foloseşte atunci cînd, prelucrînd seria de date, se obţin sporuri individuale cu baza în lanţ apropiate ca valoare unele de altele. Aceasta corespunde, în fapt, unei creşteri a nivelurilor caracteristicii studiate sub forma unei progresii aritmetice cu raţia egală cu sporul mediu.

Ajustarea pe baza sporului mediu se bazează pe relaţia care există între primul termen, sporurile cu baza în lanţ şi ultimul termen :

$$y_n = y_0 + \Delta_{1/0} + \Delta_{2/1} + \Delta_{3/2} + \ldots + \Delta_{n/n-1} \qquad (10.33)$$

Dacă se admite că abaterile în plus şi minus ale sporurilor individuale faţă de sporul mediu sînt minime şi se compensează reciproc, atunci relaţia de mai sus devine :

$$y_n = y_0 + \overline{\Delta} + \overline{\Delta} + \overline{\Delta} + \ldots + \overline{\Delta}_1$$

adică :

$$y_n = y_0 + n\overline{\Delta} \qquad (10.34)$$

Pe baza relaţiilor de mai sus se pot scrie pe rînd toţi termenii progresiei aritmetice. Pentru aceasta se consideră timpul dintre cei doi termeni extremi ai seriei ca o variabilă statistică, cu valori de la $t_1 \ldots t_n$.

[1] A se vedea capitolul „Prezentarea datelor statistice".

Scanned with OKEN Scanner

În teoria și practica statistică se obișnuiește ca termenii ajustați să se noteze cu Y pentru a-i deosebi de termenii seriei empirice care se notează cu y.

Deci, folosind acest procedeu se va obține:

$$Y_{t_i} = y_0 + t_i\bar{\Delta}, \tag{10.35}$$

în care:

y_0 reprezintă termenul considerat ca bază de ajustare;

t_i — variabila de timp luată în raport cu baza de ajustare folosită.

Urmează că, pentru a afla un termen ajustat pe baza sporului mediu, se va lua termenul de bază la care se va adăuga sporul mediu luat de un număr de unități de timp egale cu poziția pe care termenul respectiv o are față de termenul ales ca bază.

Dacă se consideră $y_0 = 17\,215$ mil. Kwh baza de ajustare și sporul mediu anual $1965 - 1975 = 3\,650,6$ mil. Kwh, în exemplul luat vor fi următoarele valori ajustate (vezi tabelul 10.3).

Valorile ajustate după sporul mediu anual înscrise în col. 2 din tabelul 10.3 s-au obținut astfel:

$$Y_{1966} = 17\,215 + 3\,650,6 = 20\,865,6;$$

$$Y_{1967} = 17\,215 + (2 \times 3\,650,6) = 24\,516,2;$$

$$Y_{1968} = 17\,215 + (3 \times 3\,650,6) = 28\,166,8 \text{ ș.a.m.d.}$$

Tabelul 10.3

Estimarea producției de energie electrică a R.S. România în perioada 1965—1975 prin sporul mediu anual

Anii	Producția de energie electrică mil. Kwh (y_i)	Variația de timp (t_i)	Valoarea teoretică a producției de energie electrică pe baza sporului mediu anual $Y_{t_i} = y_0 + t_i\bar{\Delta}$
[1965]	17 215	0	17 215
1966	20 806	1	20 865,6
1967	24 769	2	24 516,2
1968	27 828	3	28 166,8
1969	31 509	4	31 817,4
1970	35 088	5	35 468,0
1971	39 454	6	39 118,6
1972	43 439	7	42 769,2
1973	46 779	8	46 419,8
1974	49 063	9	50 070,4
1975	53 721	10	53 721

Notă: Anul închis în chenar este folosit ca bază de ajustare.

Scanned with OKEN Scanner

Obținerea termenilor ajustați prin sporul mediu se poate face calculînd succesiv termenii unei progresii aritmetice, astfel :

$$Y_{1965} = 17\,215;$$

$$Y_{1966} = 17\,215 + 3\,650{,}6 = 20\,865{,}6;$$

$$Y_{1967} = 20\,865{,}6 + 3\,650{,}6 = 24\,516{,}2;$$

$$Y_{1968} = 24\,516{,}2 + 3\,650{,}6 = 28\,166{,}8 \text{ ș.a.m.d.}$$

Reprezentînd în același grafic și valorile reale și valorile ajustate în funcție de timp se poate urmări măsura în care procedeul ales corespunde sau nu tendinței obiective de dezvoltare a fenomenului studiat.

În legătură cu ajustarea pe baza sporului mediu se precizează că nu în toate cazurile poate fi folosit ca an de ajustare anul de bază al seriei, deoarece atunci nu se ține seama decît de diferența de mărime dintre pri-

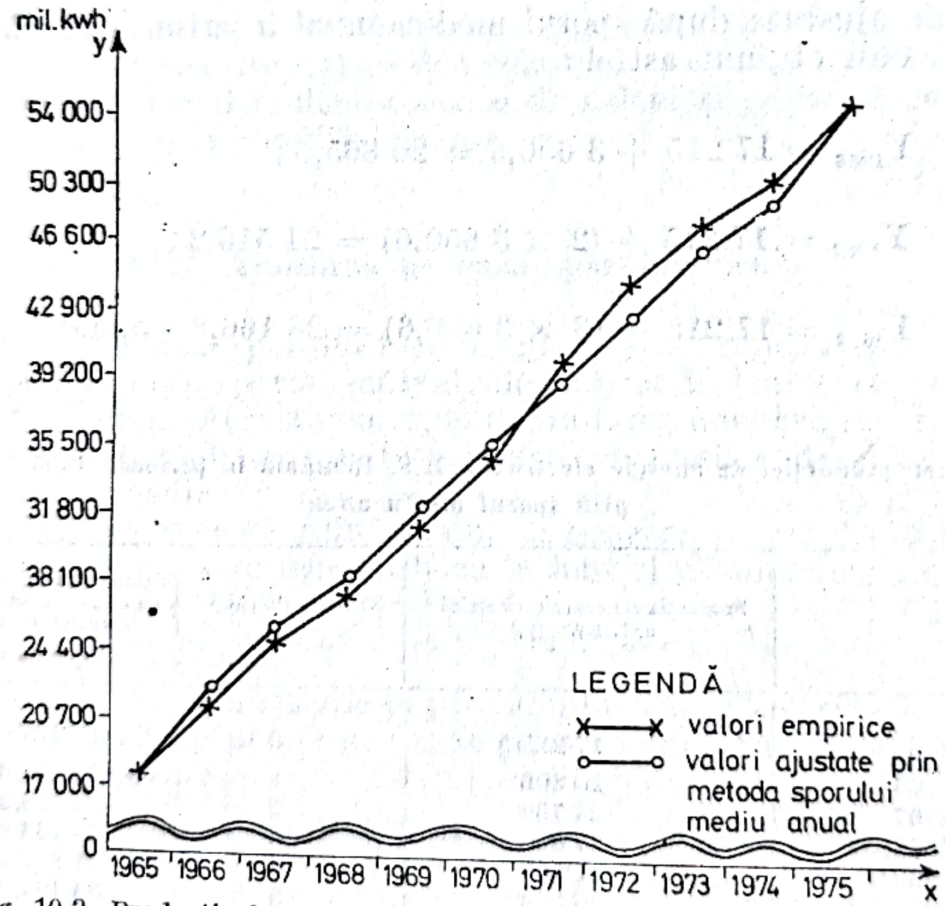


Fig. 10.2. Producția de energie electrică în R.S. România în anii 1965—1975.

mul și ultimul termen al seriei. De aceea, pentru a mări gradul de precizie al ajustării după acest procedeu, este indicat ca alegerea bazei de ajustare să se facă după ajustarea vizuală, adică se va alege din grafic acel termen care, prin poziția sa, se apropie cel mai mult de linia dreaptă teoretică ce unește cele două puncte extreme ale seriei. Se apreciază că în punctul res-

Scanned with OKEN Scanner

pectiv s-a realizat cel mai bine relaţia de progresie aritmetică dintre primul termen, sporurile anuale cu baza în lanţ şi ultimul termen. Pentru seria luată în studiu, care urmează, în general, forma de progresie aritmetică, nu este necesar să se schimbe baza de ajustare. Cînd se va alege o altă bază de ajustare se vor obţine valori de timp cu sens negativ pentru termenii care se găsesc înaintea termenului respectiv şi valori de timp pozitive pentru termenii următori.

Deci, ecuaţia de ajustare va fi :

$$Y_{t_i} = y_0 \pm t_i \overline{\Delta} \tag{10.36}$$

În consecinţă, se vor obţine două progresii aritmetice : una descrescătoare calculată de la baza de ajustare către termenii precedenţi şi alta crescătoare calculată din aceeaşi bază către termenii ultimi ai seriei.

10.5.4. Ajustarea pe baza indicelui mediu de creştere

Se foloseşte atunci cînd termenii au tendinţă de creştere sub forma unei progresii geometrice, în care raţia poate fi considerată ca egală cu indicele mediu de creştere ($\bar{I}$).

Ajustarea cu ajutorul indicelui mediu de creştere se bazează pe relaţia dintre primul termen, indicii de creştere cu baza în lanţ şi ultimul termen. Dacă ultimul termen se scrie în funcţie de primul, atunci acesta este egal cu primul termen multiplicat succesiv cu indicii cu baza în lanţ, adică :

$$y_n = y_0 \cdot I_{1/0} \cdot I_{2/1} \cdot I_{3/2} \cdots I_{n/n-1} \tag{10.37}$$

Dacă indicii cu baza în lanţ au valori apropiate între ei, atunci aceste valori pot fi substituite cu valoarea indicelui mediu de creştere. Aceasta se bazează pe relaţia dintre indicii cu baza în lanţ şi valoarea medie a lor. În relaţia precedentă se va obţine :

$$y_n = y_0 \cdot \bar{I} \cdot \bar{I} \cdot \bar{I} \ldots \bar{I}, \tag{10.38}$$

de unde :

$$y_n = y_0 \cdot \bar{I}^n. \tag{10.39}$$

Şi în acest caz se pot stabili toţi termenii progresiei geometrice calculaţi în funcţie de baza de ajustare şi raţie. Astfel un termen oarecare ajustat este egal cu termenul ales ca bază, înmulţit cu indicele mediu de creştere ridicat la o putere egală cu numărul unităţilor de timp corespunzătoare poziţiei sale faţă de termenul de bază. Şi aici, dacă termenul ales ca bază de ajustare este în interiorul seriei, atunci ecuaţiile termenilor ajustaţi sînt :

$$Y_{t_i} = y_0 \cdot \bar{I}^{\pm t_i}. \tag{10.40}$$

Scanned with OKEN Scanner

Pentru producția de energie electrică, folosind ca bază de ajustare producția anului 1965, cu ajutorul indicelui mediu anual de creștere ($\bar{I}_{1965-1975}=1{,}121$) se vor obține valorile teoretice calculate cu o progresie geometrică crescătoare.

Tabelul 10.4.

Estimarea producției de energie electrică pe perioada 1965—1975, folosind procedeul indicelui mediu de creștere

Anii	Producția de energie electrică (mil.Kwh) y_t	Variația de timp t_i	Valorile teoretice ale producției de energie electrică $Y_{t_i} = 17\,215 \cdot 1{,}1205^{t_i}$
0	1	2	3
[1965]	17 215	0	17 215
1966	20 806	1	19 289,4
1967	24 769	2	21 613,8
1968	27 828	3	24 218,3
1969	31 509	4	27 136,6
1970	35 088	5	30 406,6
1971	39 454	6	34 070,6
1972	43 459	7	38 176,1
1973	46 779	8	42 776,3
1974	49 063	9	47 930,8
1975	53 721	10	53 706,5

Notă: Anul inclus în chenar este folosit ca bază de ajustare.

Valorile ajustate după indicele mediu de creștere (col. 3, tabelul 10.4) se pot obține în două moduri:

$$Y_{1965} = 17\,215;$$

$$Y_{1966} = 17\,215 \cdot 1{,}1205 = 19\,289{,}4;$$

$$Y_{1967} = 17\,215 \cdot 1{,}1205^2 = 21\,613{,}8 \text{ ș.a.m.d.}$$

sau ca progresie geometrică:

$$Y_{1965} = 17\,215;$$

$$Y_{1966} = 17\,215 \cdot 1{,}1205 = 19\,289{,}4;$$

$$Y_{1967} = 19\,289{,}4 \cdot 1{,}1205 = 21\,613{,}8 \text{ ș.a.m.d.}$$

Pentru a verifica dacă procedeul ales corespunde legității obiective de dezvoltare a producției de energie electrică în perioada analizată se

Scanned with OKEN Scanner

reprezintă în același grafic atît valorile empirice cît și cele ajustate după indicele mediu de creștere, luate în funcție de timp :

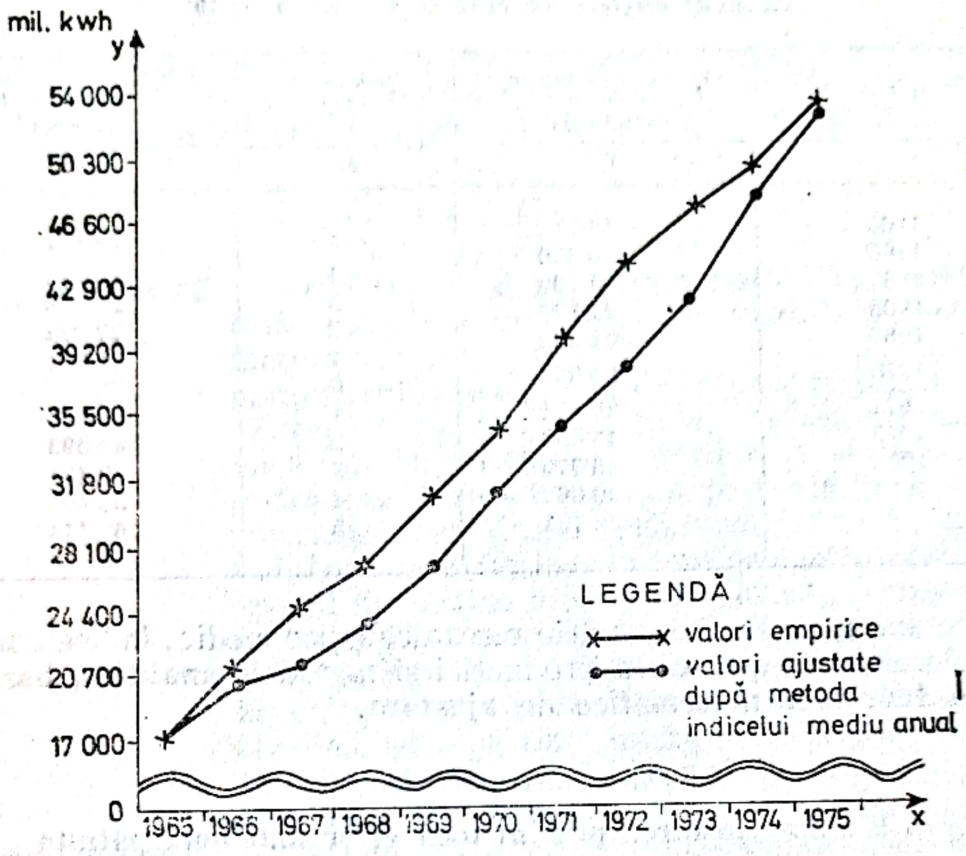


Fig. 10.3. Producția de energie electrică în R.S. România în anii 1965—1975.

După cum se observă din tabel, și mai ales din grafic, valorile ajustate după acest procedeu nu corespund formei obiective de creștere a producției de energie electrică a României pe perioada 1965—1975.

Pentru a verifica care este subperioada în care indicele mediu calculat este semnificativ, se va folosi și altă bază de ajustare situată în interiorul seriei. Astfel, luînd în considerare valorile indicilor cu baza în lanț și forma tendinței desprinsă din grafic, se constată că în anul 1971/1970 apare cea mai mică abatere față de ritmul mediu anual de creștere. Deci se poate folosi pentru completarea analizei anul 1970, ca bază de ajustare (tab. 10.5).

Comparînd cele două serii de date ajustate obținute prin același procedeu, dar cu bază diferită de ajustare (1965 și 1970), se poate observa că, folosind ca bază de ajustare producția anului 1970, se estimează mai corect valorile teoretice care ar corespunde creșterii producției de energie electrică după o progresie geometrică. Mai mult, comparînd datele se poate constata că deși pe toată seria estimarea producției de forma unei progresii geometrice cu rația egală cu indicele mediu nu este corespunzătoare, totuși între anii 1967—1972 această tendință se verifică. În consecință, se poate admite ipoteza că seria prezintă două tendințe de variație de scurtă durată : una de formă liniară, specifică termenilor extremi și alta de formă exponențială, specifică termenilor mediani. Pentru verificare

Tabelul 10.5

Estimarea producţiei de energie electrică pe perioada 1965—1975 folosind procedeul indicelui mediu de creştere, cu bază de ajustare anul 1970

Anii	Producţia de energie electrică mil.Kwh (y_t)	Variaţia de timp (t_i)	Valorile teoretice ale producţiei de energie electrică $Y_{t_i} = y_0 \bar{I} \pm t_i$
0	1	2	3
1965	17 215	−5	19 821
1966	20 806	−4	22 220
1967	24 769	−3	24 908
1968	27 829	−2	27 922
1969	31 509	−1	31 301
1970	35 088	0	35 088
1971	39 454	1	39 333
1972	43 439	2	44 093
1973	46 779	3	49 428
1974	49 063	4	55 409
1975	53 721	5	62 114

este necesar ca, pe lîngă metodele mecanice (spor mediu, indicele mediu), să se folosească la estimarea producţiei şi metodele analitice, bazate pe folosirea funcţiilor matematice de ajustare.

10.5.5. Ajustarea prin metoda celor mai mici pătrate

Ajustarea seriilor cronologice prin procedeele prezentate pînă aici se bazează doar pe relaţia care există între primul şi ultimul termen al seriei luată în funcţie de numărul termenilor şi exprimată sub formă de mărime absolută — sporul mediu — sau sub formă de mărime relativă — indicele mediu de creştere. De aceea, aceste procedee sînt cunoscute în literatura de specialitate şi sub denumirea de procedee mecanice. Ori, asupra formei de variaţie influenţează şi termenii intermediari ai unei serii dinamice, care pot prezenta abateri de la modificarea sistematică produsă de legitatea care le generează. În acest sens, seria cronologică poate fi considerată ca o variabilă în timp, care se formează ca o funcţie de forma :

$$y = f(t_i),$$

în care :

t reprezintă valorile variabilei independente (timpul);

y — valorile variabilei dependente (fenomenele care sînt prezentate în seria cronologică).

În practica statistică, cel mai frecvent se întîlnesc fenomenele care, în dinamica lor, apar sub forma unor funcţii liniare, exponenţiale, parabolice sau a unei hiperbole. În continuare se va prezenta ajustarea pe baza funcţiilor liniare, exponenţiale şi parabolice, ca fiind cele mai des utilizate în ajustarea seriilor cronologice.

Scanned with OKEN Scanner

Alegerea funcţiei care corespunde cel mai bine formei reale de evoluţie a fenomenelor se face tot pe baza unei analize atente a graficului şi indicatorilor absoluţi şi relativi ce caracterizează seria empirică de date, după cum rezultă în continuare.

În cazul cînd graficul arată o tendinţă de creştere absolută constantă, verificată şi printr-o variaţie mică a sporurilor cu baza în lanţ, se poate aprecia că fenomenul creşte după o funcţie liniară a cărei ecuaţie de estimare este :

$$Y_{t_i} = a + bt_i, \tag{10.41}$$

în care :

Y_{t_i} reprezintă valorile teoretice ale caracteristicii ce trebuie ajustată, calculată în funcţie de valorile caracteristicii factoriale (t) ;

a — parametrul care are sens de mărime medie şi arată la ce valoare ar fi ajuns caracteristica rezultativă y, dacă influenţa tuturor factorilor, în afara celui înregistrat, ar fi fost constantă pe toată perioada ;

b — parametrul care sintetizează numai influenţa caracteristicii factoriale (t) şi care în sens geometric arată tocmai gradul de înclinaţie a liniei (curbei) de regresie. De menţionat că în ecuaţiile de regresie a unei serii cronologice b nu poate să fie decît pozitiv în cazul seriilor cu tendinţă obiectivă de creştere continuă şi negativ cînd seria prezintă tendinţa de scădere continuă ;

t_i — valorile caracteristicii factoriale care, în cazul seriilor dinamice, este timpul.

Cînd graficul arată o tendinţă de creştere relativ, constantă, respectiv creşteri absolute din ce în ce mai mari, verificată şi prin obţinerea unor valori apropiate ale indicilor cu baza în lanţ, atunci se poate aprecia că fenomenul creşte de forma unei *funcţii exponenţiale* a cărei ecuaţie de estimare este :

$$Y_{t_i} = ab^{t_i}. \tag{10.42}$$

Cînd pe grafic se obţine o curbă care are fie un punct de maxim, fie un punct de minim, atunci se apreciază că fenomenul studiat se modifică în timp sub forma unei *parabole de gradul doi*. Se pot întîlni cazuri frecvente cînd pe grafic apare numai o secţiune de parabolă crescătoare sau descrescătoare specifică numai anumitor etape. În acest caz este necesar să se aprecieze, prin compararea cu altă etapă precedentă, care este tendinţa de lungă durată şi să se ajusteze seria în consecinţă.

Ecuaţia de estimare a unei parabole de gradul doi exprimată în funcţie de timp este :

$$Y_{t_i} = a + bt_i + ct_i^2. \tag{10.43}$$

Ca şi în cazul corelaţiei, pentru aflarea parametrilor funcţiei de regresie necesară ajustării seriei se aplică metoda celor mai mici pătrate, adică :

$$\Sigma(y_i - Y_{t_i})^2 = \min.$$

Scanned with OKEN Scanner

În cazul funcției liniare această condiție devine :

$$\Sigma[(y_i - (a + bt_i)]^2 = \min.$$

Pentru a putea satisface această condiție este necesar să se determine valorile celor doi parametri a și b. Pentru aceasta se va folosi sistemul de ecuații normale[1], care măsoară legătura liniară dintre variabila independentă x și variabila dependentă y. Înlocuind pe x_i cu t_i se obține :

$$\begin{aligned} na + b\ \Sigma t &= \Sigma y\ ; \\ a\ \Sigma t + b\Sigma t^2 &= \Sigma ty. \end{aligned} \qquad (10.44)$$

Analizînd însă gradul de dependență dintre caracteristica de timp t și volumul producției pentru care trebuie să se calculeze valorile ajustate Y, se poate constata că producția nu este în funcție de timp. Volumul producției de energie electrică — ca, în general, orice fenomen social-economic — depinde de o serie de alți factori a căror influență este prezentă în toate unitățile de timp. Astfel, producția depinde, printre altele, de gradul de înzestrare tehnică, de gradul de calificare a muncii, de modul de organizare a procesului de muncă etc. Din acest motiv, pentru a anihila influența variației de timp, se pune condiția ca :

$$\Sigma t_i = 0. \qquad (10.45)$$

Pentru $\Sigma t_i = 0$, sistemul de ecuații normale prezentat anterior devine :

$$\begin{aligned} na &= \Sigma y, \\ b\Sigma t^2 &= \Sigma ty \end{aligned} \quad \text{de unde} \quad \begin{aligned} a &= \frac{\Sigma y_i}{n} \\ b &= \frac{\Sigma ty_i}{\Sigma t_i^2} \end{aligned} \qquad (10.46)$$

Și pe baza rezolvării sistemului de ecuații se demonstrează că valoarea lui a este egală chiar cu media seriei, adică valoarea formată numai pe baza influenței constante a cauzelor esențiale, deoarece :

$$a = \frac{\Sigma y_i}{n} = \bar{y} \qquad (10.47)$$

Deci, condiția aplicării acestui procedeu de ajustare este $\Sigma t_i = 0$. Pentru a satisface această condiție trebuie să se considere originea valorilor de timp ca fiind în centrul seriei. Astfel, seria va fi divizată în două părți, în care numărul unităților de timp negative trebuie să fie egal cu numărul celor pozitive. În cazul cînd seria este

[1] Vezi capitolul „Analiza legăturilor dintre fenomenele social-economice".

formată dintr-un număr impar de termeni, originea valorilor de timp va fi chiar în dreptul termenului central și variația de timp se va măsura în intervale întregi : 0 ; ±1 ; ±2 ; ±3 etc.

În cazul unei serii dinamice formate dintr-un număr par de termeni, originea valorilor de timp va cădea între cei doi termeni centrali și variația de timp se va măsura în jumătăți de intervale de timp : 0 ; ±1 ; ±3 ; ±5 etc.

În exemplul prezentat, avînd 11 termeni, centrul seriei este anul 1970, care se notează cu zero. În plus, este necesar să se calculeze valorile lui t_i^2 și $t_i y_i$ pentru a se rezolva sistemul de ecuații normale, necesare obținerii ecuațiilor de regresie corespunzătoare fiecărei valori a lui t.

Tabelul 10.6

Ajustarea producției de energie electrică pe perioada 1965—1975, pe baza funcției liniare

Anii	Producția de energie electrică	t_i	t_i^2	$t_i y_i$	$Y_{t_i} = 35\,424{,}6 + 3\,643\,t_i$
0	1	2	3	4	5
1965	17 125	−5	25	−86 075	17 109,6
1966	20 806	−4	16	−83 224	20 752,6
1967	24 769	−3	9	−74 307	24 395,6
1968	27 828	−2	4	−55 656	28 138,6
1969	31 509	−1	1	−31 509	31 781,6
1970	35 088	0	0	0	35 424,6
1971	39 454	1	1	39 454	39 067,6
1972	43 439	2	4	86 878	42 710,6
1973	46 779	3	9	140 337	46 353,6
1974	49 063	4	16	196 252	49 996,6
1975	53 721	5	25	268 605	53 639,6
Total	389 671	—	110	−330 771	$\Sigma Y_i = \Sigma y_i$
	Σy_i	$\Sigma t_i = 0$	Σt_i^2	731 526	
				400 755	

Pe baza datelor calculate în tabel se va obține următorul sistem de rezolvat :

$$11\,a = 389\,671 \qquad a = \frac{389\,671}{11} = 35\,424{,}6$$

unde

$$110\,b = 400\,755 \qquad b = \frac{400\,755}{110} = 3\,643$$

Deci :

$$Y_{t_i} = 35\,424{,}6 + 3\,643\ t_i.$$

Pe ani, ecuațiile de regresie corespunzătoare fiecărei valori ale variabilei t vor fi :

$$Y_{1965} = 35\,424{,}6 + 3\,643\ (-5) = 17\,109{,}6\,;$$

$$Y_{1966} = 35\,424{,}6 + 3\,643\ (-4) = 20\,752{,}6.$$

Scanned with OKEN Scanner

Reprezentînd grafic se poate observa relaţia dintre valorile reale şi cele ajustate.

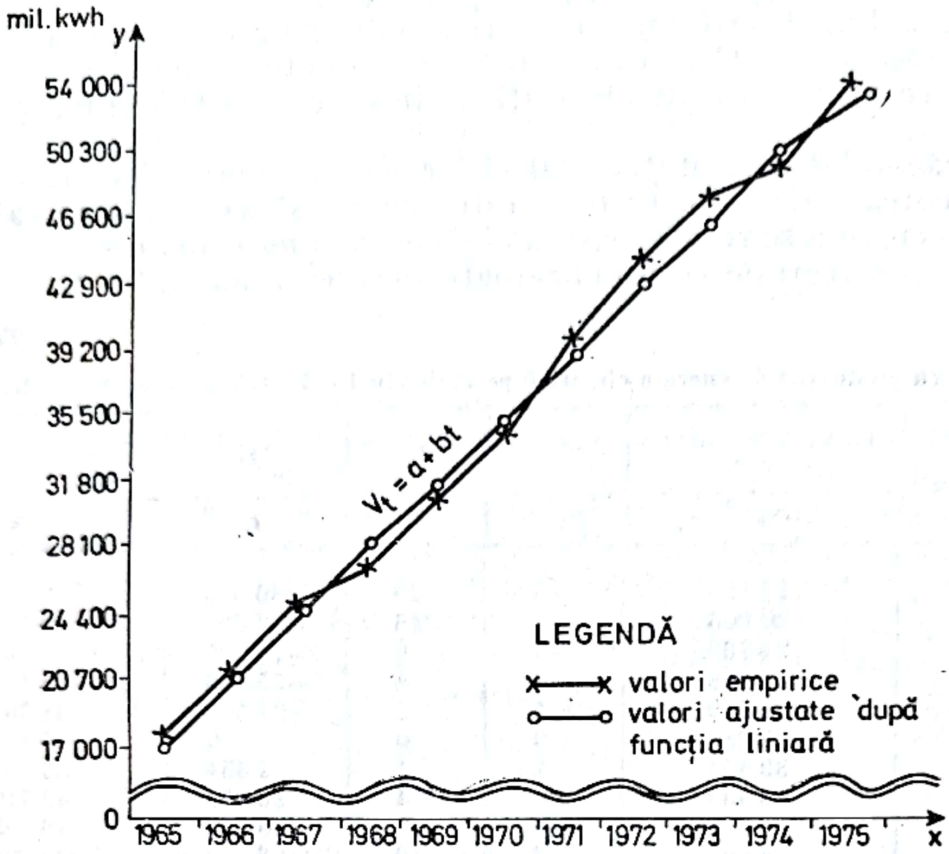


Fig. 10.4. Producţia de energie electrică în R.S. România în anii 1965—1975.

Verificarea calculării ecuaţiilor de regresie se face prin însumarea lor, operaţie prin care se obţine o valoare aproximativ egală cu suma valorilor empirice ale termenilor seriei. Aceasta se bazează pe faptul că prin ajustare s-au redistribuit influenţele factorilor astfel : toţi factorii au fost consideraţi cu influenţă constantă pe toată perioada şi variabil a fost numai timpul.

În cazul ajustării pe baza funcţiei exponenţiale condiţia de minim :

$$\Sigma(y_i - Y_{t_i})^2 = \min$$

devine :

$$\Sigma[y_i - (ab^{t_i})]^2 = \min.$$

Fiind o funcţie exponenţială, ecuaţia ei va fi :

$$\log y_i = \log a + t_i \log b \qquad (10.48)$$

Sistemul de ecuaţii normale, prezentat la capitolul „Analiza legăturilor dintre fenomenele social-economice", pentru legătura de formă exponenţială, se adaptează pentru cazul în care variabila factorială x este con-

Scanned with OKEN Scanner

siderată variabila de timp t, adică:

$$n \log a + \log b\Sigma t_i = \Sigma \log y_i \qquad (10.49)$$

$$\log a\ \Sigma t_i + \log b\Sigma t_i^2 = \Sigma(t_i \log y_i)$$

care, pentru aceeaşi condiţie pusă la seriile dinamice, ca $\Sigma t_i = 0$, sistemul devine:

$$n \log a = \Sigma \log y_i$$

$$\log b\ \Sigma t_i^2 = \Sigma\ (t_i \log y_i). \qquad (10.50)$$

Se poate observa că, prin logaritmare, sistemul de ecuaţii normale este asemănător cu cel obţinut pentru funcţia liniară, cu deosebirea că aici se folosesc logaritmii, iar valorile celor doi parametri a şi b se obţin prin antilogaritmare.

Aplicînd aceste procedee la exemplul precedent, se vor înscrie datele într-un tabel: (vezi tabelul 10.7).

Deci, sistemul de ecuaţii care trebuie rezolvat pentru aflarea valorii logaritmilor celor doi parametri va fi:

$$11 \log a = 49{,}77096;$$

$$110 \log b = 5{,}31739,$$

de unde:

$$\log a = \frac{49{,}77096}{11} = 4{,}52372;\ \text{antilog} = 33\,400;$$

$$\log b = \frac{5{,}31739}{110} = 0{,}04834;\ \text{antilog} = 1{,}118.$$

Deci, log ecuaţiei funcţiei de ajustare va fi:

$$\log Y_{t_i} = 4{,}52372 + 0{,}04834\ t_i.$$

Prin antilogaritmare se obţine ecuaţia funcţiei exponenţiale necesară ajustării seriei privind producţia de energie electrică a R.S. România pe perioada 1965—1975.

$$Y_{t_i} = 33\,400 + 1{,}118^{t_i}$$

Şi în acest caz, pentru verificarea obiectivităţii funcţiei folosite se vor reprezenta grafic seriile dinamice ale valorilor reale şi ale valorilor teoretice calculate pe baza ecuaţiilor de regresie de tip exponenţial.

Scanned with OKEN Scanner

Tabelul 10.7

Ajustarea producției de energie electrică pe perioada 1965—1975, pe baza funcției exponențiale

Anii	Producția de energie electrică (y_t)	t_i	t_i^2	$\log y_t$	$t_i (\log y_t)$	$\log Y_{t_i} = \log a + t_i \log b$	$Y_{t_i} = ab^{t_i}$
0	1	2	3	4	5	6	7
1965	17 215	—5	25	4,23591	—21,17955	4,28202	19 150
1966	20 806	—4	16	4,31818	—17,27272	4,33036	21 400
1967	24 769	—3	9	4,39270	—13,17810	4,37870	23 920
1968	27 828	—2	4	4,44451	— 8,88902	4,42704	26 730
1969	31 509	—1	1	4,49845	— 4,49845	4,47538	29 880
1970	35 088	0	0	4,54642	0	4,52372	33 400
1971	39 454	1	1	4,59605	4,59605	4,57206	37 330
1972	43 439	2	4	4,63789	9,27578	4,62040	41 730
1973	46 779	3	9	4,68006	14,04018	4,66874	46 640
1974	49 063	4	16	4,69073	18,76292	4,71708	52 130
1975	53 721	5	25	4,73006	23,65030	4,76542	58 130
1965—1975	389 671	0	110	49,77096	70,32523 —65,01784 5,31739	49,76092	390 440
	Σy_t	Σt_i	Σt_i^2	$\Sigma \log y_t$	$\Sigma t_i (\log y_t)$		

Scanned with OKEN Scanner

În cazul în care ajustarea seriei se face după ecuaţiile unei funcţii de forma unei parabole de gradul 2, se obţine următorul sistem de ecuaţii normale adaptate la factorul timp :

$$\begin{aligned} na + b\ \Sigma t_i + c\ \Sigma t_i^2 &= \Sigma y_i \\ a\Sigma t_i + b\ \Sigma t_i^2 + c\ \Sigma t_i^3 &= \Sigma t_i y_i \\ a\ \Sigma t_i^2 + b\ \Sigma t_i^3 + c\ \Sigma t_i^4 &= \Sigma t_i^2 y_i \end{aligned} \qquad (10.51)$$

Punînd condiţia ca $\Sigma t_i = 0$ şi $\Sigma t_i^3 = 0$, sistemul de ecuaţii normale ce trebuie rezolvate este :

$$\begin{aligned} na + c\ \Sigma t_i^2 &= \Sigma t_i y_i \\ b\ \Sigma t_i^2 &= \Sigma\ t_i\ y_i \\ a\ \Sigma t_i^2 + c\ \Sigma t_i^4 &= \Sigma t_i^2 y_i \end{aligned} \qquad (10.52)$$

Din ecuaţia a doua se calculează valoarea lui b care este aceeaşi ca şi în ecuaţia liniei drepte (10.45).

Din ecuaţiile 1 şi 3 se calculează valorile parametrilor a şi c.

$$\begin{aligned} na + c\ \Sigma t_i^2 &= \Sigma y_i \\ a\ \Sigma t_i^2 + c\ \Sigma t_i^4 &= \Sigma t_i^2 y_i \end{aligned}$$

Pentru rezolvarea sistemului se folosesc datele din tabelul 10.8.

Tabelul 10.8

Ajustarea producţiei de energie electrică pe perioada 1965—1975, pe baza funcţiei parabolice de gradul 2

Anii	Producţia de energie electrică (mil. Kwh) y_i	t_i	t_i^2	t_i^4	$t_i y_i$	$t_i^2 y_i$	$Y_{t_i} = a + bt_i + ct_i^2$
A	1	2	3	4	5	6	7
1965	17 215	−5	25	625	−86 075	430 375	17 113,5
1966	20 806	−4	16	256	−83 224	332 896	20 813,8
1967	24 769	−3	9	81	−74 307	222 921	24 501,3
1968	27 828	−2	4	16	−55 656	111 312	28 176,2
1969	31 509	−1	1	1	−31 509	31 509	31 838,5
1970	35 088	0	0	0	0	0	35 488,0
1971	39 454	1	1	1	39 454	39 454	39 137,5
1972	43 439	2	4	16	86 878	173 756	42 799,8
1973	46 779	3	9	81	140 337	421 011	46 474,7
1974	49 063	4	16	256	196 252	785 008	49 972,0
1975	53 721	5	25	625	268 605	1 343 025	53 545,5
—	389 671	0	110	1958	−390 771	3 891 267	389 820,9
					731 526		
					400 755		
	Σy_i	Σt_i	Σt_i^2	Σt_i^4	$\Sigma t_i y_i$	$\Sigma t_i^2 y_i$	ΣY_{t_i}

Scanned with OKEN Scanner

Pentru determinarea parametrilor ecuației medii de estimare, de forma unei parabole de gradul 2, sistemul de ecuații normale va fi :

$$11a + 110\ c = 389\ 671$$

$$110\ b = 400\ 755$$

$$110\ a + 1958\ t^4 = 3\ 891\ 267$$

Rezolvînd sistemul se obține : $a = 35\ 488$; $b = 3\ 643{,}2$; $c = -6{,}34$.

Deci, ecuația medie de estimare a producției de energie electrică pe perioada 1965—1975, după o funcție parabolică de gradul 2, este :

$$Y_{1965-1975} = 35\ 488 + 3\ 643{,}2\ t_i - 6{,}34\ t_i^2.$$

Pe ani, ecuațiile calculate în funcție de valoarea lui t_i sînt înscrise în tabelul 10.8, col. 7.

Și în acest caz, reprezentînd grafic datele empirice și cele teoretice, se poate stabili vizual gradul de abatere dintre cele două curbe.

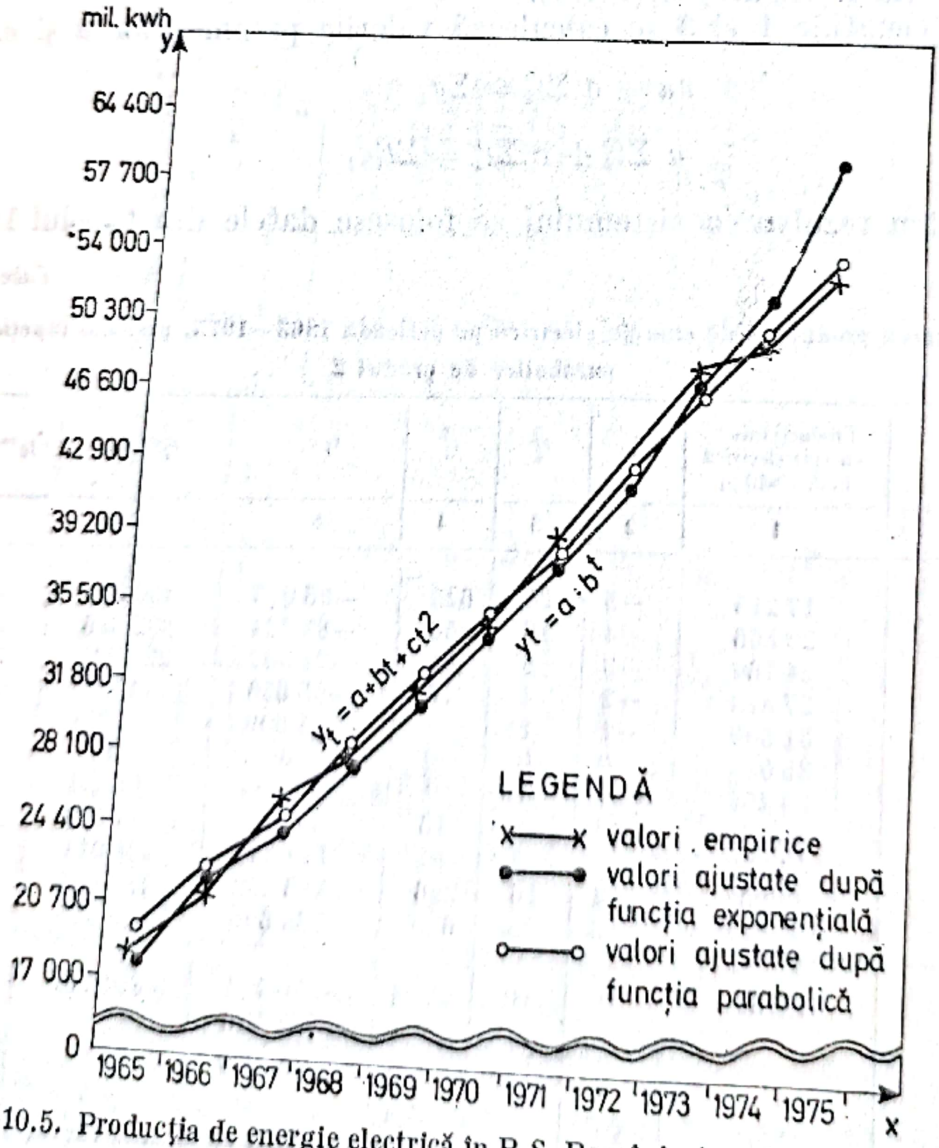


Fig. 10.5. Producția de energie electrică în R.S. România în anii 1965—1975.

10.5.6. Criterii de alegere a procedeelor de ajustare

Folosind mai multe procedee de ajustare a seriei privind producţia de energie electrică pe perioada 1965—1975, s-au obţinut mai multe valori teoretice pentru acelaşi an. Pentru a alege procedeul cel mai adecvat, în acest caz, este necesar să se compare rezultatele obţinute.

Compararea rezultatelor se face folosind valorile parametrilor funcţiilor de ajustare, reprezentarea grafică şi coeficienţii de variaţie a valorilor empirice de la valorile ajustate.

În primul rînd, se poate observa că ajustînd seria după cele trei funcţii analitice, valoarea medie a variabilei y în funcţie de timp este :

— după funcţia rectilinie $a = 35\,424{,}6$ mil. Kwh ;

— după funcţia exponenţială $a = 33\,400$ mil. Kwh ;

— după funcţia parabolică de gradul 2 $a = 35\,488$ mil. Kwh.

Faţă de valoarea mediană a termenilor reali, cea mai mică abatere o are valoarea medie estimată prin ecuaţia liniei drepte (+336,6 mil. Kwh) urmată de aceea estimată prin parabola de gradul 2(+400 mil. Kwh) şi cea mai mare abatere şi de semn invers este aceea prin ecuaţia exponenţială (—1 688 mil. Kwh).

Dacă se compară parametrul b al liniei de regresie cu sporul mediu se constată că există o abatere foarte mică. Sporul mediu calculat pe baza sporului cu baza în lanţ ($\bar{\Delta}$) este egal cu 3 650,6, iar cel estimat prin linia dreaptă $b = 3\,643$, deci o abatere nesemnificativă, de 7 unităţi. Comparînd însă indicele mediu de creştere calculat din datele empirice $\bar{I}$, folosit la ajustarea prin procedeele mecanice, cu cel estimat prin funcţia exponenţială (b), se constată o abatere mai semnificativă $\bar{I} = 1{,}1205$, iar $b = 1{,}118$ din ecuaţia $Y_{t_i} = ab^{t_i}$, şi această comparaţie conduce la acceptarea ipotezei că în perioada luată în calcul producţia de energie electrică se apropie mai mult de o funcţie liniară pe toată perioada şi cu tendinţe de creştere exponenţială pe subperioade. Din această cauză este de aşteptat ca cele mai mici abateri să se obţină faţă de valorile estimate prin parabola de gradul 2.

În al doilea rînd, comparînd pe grafic datele reale cu cele ajustate se constată că în perioada 1965—1975 producţia de energie electrică s-a dezvoltat de forma unei parabole de gradul 2, deoarece această linie de tendinţă prezintă cele mai mici abateri.

În al treilea rînd, pentru a aprecia cu mai multă siguranţă care este forma de ajustare cea mai potrivită seriei de date empirice se calculează suma abaterilor luate în valoarea absolută dintre datele empirice şi cele ajustate $|y_i - Y_{t_i}|$[1]. Se consideră cel mai potrivit procedeu acela la care se obţine $\Sigma|y - Y_{t_i}| = \min$. Din datele tabelului 10.9 rezultă că, folosind acest criteriu, funcţia de ajustare cea mai potrivită este parabola de gradul 2.

În al patrulea rînd, pentru a se putea aprecia obiectivitatea procedeului ales, se poate calcula coeficientul de variaţie ca raport între abaterea

---

[1] Se poate folosi pentru comparaţie şi suma pătratelor abaterilor $\Sigma(y_i - Y_{t_i})^2$ dar este o operaţie mai complicată, care nu modifică raportul de variaţie de la un tip de procedeu la altul.

medie liniară de la valorile ajustate și nivelul mediu al termenilor seriei empirice :

$$V'_{y/t} = \frac{\bar{d}_{y/t}}{\bar{y}} \tag{10.53}$$

în care :

$\bar{d}_{y/t}$ reprezintă abaterea medie liniară a valorilor reale de la valorile ajustate, calculată după formula :

$$\bar{d}_{y/t} = \frac{\Sigma |d_{y/t}|}{n} = \frac{\Sigma |y_t - Y_{t_t}|}{n}. \tag{10.54}$$

Deci, pentru exemplul considerat se vor calcula cinci abateri medii liniare :

$$\bar{d}_{y/t} \begin{cases} = \dfrac{\Sigma[|y - (y_0 + t\overline{\Delta})|]}{n} = \dfrac{3\,711{,}8}{11} = 337{,}4 \\ = \dfrac{\Sigma[|y - (y_0 \cdot \bar{I}^t)|]}{n} = \dfrac{3\,3131}{11} = 3\,011{,}9 \\ = \dfrac{\Sigma[|y - (a + bt)|]}{n} = \dfrac{3\,907{,}2}{11} = 355{,}4 \\ = \dfrac{\Sigma[|y - (ab^t)|]}{n} = \dfrac{19\,241}{11} = 1\,749{,}2 \\ = \dfrac{\Sigma[|y - (a + bt + ct^2)|]}{n} = \dfrac{3\,589{,}7}{11} = 326{,}3 \end{cases}$$

Pe baza abaterilor medii liniare și a valorii medii a termenilor seriei se vor obține următorii coeficienți de variație, pentru :

— ajustarea după sporul mediu :

$$V'_{y/t} = \frac{\bar{d}_{y/t}}{\bar{y}} \cdot 100 = \frac{337{,}4}{35\,424{,}6} \cdot 100 = 0{,}95\%;$$

— ajustarea după indicele mediu de creștere :

$$V'_{y/t} = \frac{\bar{d}_{y/t}}{\bar{y}} \cdot 100 = \frac{3\,011{,}9}{35\,424{,}6} \cdot 100 = 8{,}5\%;$$

— ajustarea pe baza funcției rectilinii :

$$V'_{y/t} = \frac{\bar{d}_{y/t}}{\bar{y}} \cdot 100 = \frac{355{,}4}{35\,424{,}6} \cdot 100 = 1{,}00\%;$$

Scanned with OKEN Scanner

Tabelul 10.9

Variația valorilor empirice de la valorile ajustate

Anii	Datele inițiale	Date ajustate și abaterile (sub formă de modul) termenilor inițiali de la cei ajustați după procedeele:									
		Sporului mediu		Indicelui mediu		Ecuația liniei drepte		Ecuația funcției exponențiale		Ecuația unei parabole de gradul 2	
		$Yt_i = y_0 + t_i\overline{\Delta}$	$\lvert y_i - Yt_i \rvert$	$Yt_i = y_0\overline{I}^{t_i}$	$\lvert y_i - Yt_i \rvert$	$Yt_i = a + bt_i$	$\lvert y_i - Yt_i \rvert$	$Yt_i = ab^{t_i}$	$\lvert y_i - Yt_i \rvert$	$Yt_i = a + bt_i + ct_i^2$	$\lvert y_i - Yt_i \rvert$
0	1	2	3	4	5	6	7	8	9	10	11
1965	17 215	17 215	0	17 215	0	17 109,6	105,4	19 150	1 935	17 113,5	101,5
1966	20 806	20 865,6	59,6	19 289,4	1 516,0	20 752,6	53,4	21 400	594	20 813,8	7,8
1967	24 769	24 516,2	252,8	21 613,8	3 155,2	24 395,6	283,4	23 920	849	24 501,3	267,7
1968	27 828	28 166,8	338,8	24 218,3	3 609,7	28 138,6	310,6	26 730	1 098	28 176,2	348,2
1969	31 509	31 817,4	308,4	27 136,6	4 372,4	31 781,6	172,6	29 880	1 629	31 838,5	329,5
1970	35 088	35 468,0	380,0	30 406,6	4 681,4	35 424,6	336,6	33 400	1 688	35 488,0	366,0
1971	39 454	39 118,6	335,4	34 070,6	5 383,4	39 067,6	386,4	37 330	2 124	39 137,5	316,5
1972	43 439	42 769,2	669,8	38 176,1	5 262,9	42 710,6	728,4	41 730	1 709	42 799,8	639,2
1973	46 779	46 419,8	359,2	42 776,3	4 002,7	46 353,6	425,4	46 640	139	46 474,7	304,3
1974	49 063	50 070,4	1 007,4	47 930,8	1 132,2	49 966,6	933,6	52 130	3 067	49 972,0	909,0
1975	53 721	53 721	0	53 706,3	14,5	53 639,6	81,4	58 130	4 409	53 545,5	175,5
	389 671	—	3 711,8	—	33 131,0	—	3 907,2	—	19 241	—	3 589,7

Scanned with OKEN Scanner

— ajustarea pe baza funcției exponențiale :

$$V'_{y/t} = \frac{\bar{d}_{y/t}}{\bar{y}} \cdot 100 = \frac{1\,749{,}2}{35\,424{,}6} \cdot 100 = 4{,}9\,\%;$$

— ajustarea pe baza funcției parabolice de gradul 2 :

$$V'_{y/t} = \frac{\bar{d}_{y/t}}{\bar{y}} \cdot 100 = \frac{326{,}3}{35\,424{,}6} \cdot 100 = 0{,}92\,\%.$$

Ca orice coeficient de variație, se interpretează ca fiind cel mai bun acela care are valoarea cea mai mică. În cazul ajustării seriilor cronologice se va aprecia că cele mai corespunzătoare valori teoretice sînt cele față de care se obțin cele mai mici abateri, deci, coeficientul cel mai mic de variație[1].

Interpretînd toate rezultatele, se poate trage concluzia că producția de energie electrică a R.S. România pe perioada 1965—1975 a avut tendință de creștere sub forma unei parabole de gradul 2, apropiindu-se mai mult de funcția rectiliniară. Aceasta este confirmată și de valoarea mică a coeficientului de variație, calculat ca raport între abaterea medie a termenilor individuali de la linia teoretică și nivelul mediu al seriei.

În consecință se poate spune că ajustarea termenilor unei serii cronologice nu se poate face la întîmplare, ci, interpretînd toate rezultatele, inclusiv reprezentarea grafică, se va alege acel procedeu de ajustare care ține seama de tendința obiectivă de dezvoltare din etapa analizată. De reținut este faptul că atunci cînd din grafic rezultă clar tendința de dezvoltare, nu mai este necesar să se folosească mai multe procedee de ajustare.

10.6. MĂSURAREA OSCILAȚIILOR SEZONIERE

În manifestarea concretă a fenomenelor economice, variațiile mari ale acestora nu sînt întotdeauna provocate de influența unor cauze întîmplătoare. Uneori aceste variații au un caracter de regularitate și pot fi rezultatul acțiunii unor cauze obiective, legate de modificarea anotimpurilor. Astfel de oscilații se pot întîlni în unele ramuri ale producției și circulației de mărfuri, în transporturi maritime și fluviale, în turism și altele.

Pentru a măsura influența valului sezonier este necesar să se cunoască periodicitatea producerii variației în funcție de care se vor înregistra datele statistice lunare sau trimestriale.

---

[1] La calculul coeficienților de variație a valorilor empirice de la cele teoretice se poate folosi și abaterea medie pătratică, adică :

$$V_{y/t} = \frac{\sigma_{y/t}}{\bar{y}} \cdot 100 = \sqrt{\frac{\Sigma (y_i - Yt_i)^2}{n}} \cdot 100.$$

De exemplu, în producţia şi circulaţia de mărfuri de larg consum pot fi produse care prezintă grade de sezonalitate diferite; astfel, producţia şi desfacerea de zarzavaturi poate varia foarte mult de la o lună la alta, în timp ce producţia de bere poate avea abateri foarte mici de la un trimestru la altul şi totuşi, desfacerea ei către consumatori să prezinte oscilaţii foarte mari cu caracter sezonier.

Cunoaşterea gradului de sezonalitate este necesară, în special, cînd datele se referă la o economie planificată. Determinarea gradului de sezonalitate serveşte pentru fundamentarea planului de producţie (desfacere), de aprovizionare şi al forţei de muncă. Urmărirea gradului de îndeplinire a planului se face nu numai pe ani calendaristici, care prezintă, de regulă, variaţii mult mai mici de la un an la altul, ci şi pe luni şi trimestre, unde pot apărea variaţii foarte mari. Pentru a măsura sezonalitatea trebuie, deci, ca indicatorul totalizat pe întregul an să fie prezentat pe luni sau trimestre, după cum se manifestă în mod concret sezonalitatea.

În teoria şi practica statistică, pentru măsurarea sezonalităţii anumitor fenomene se determină indicele (coeficientul) de sezonalitate. Mai frecvent aceasta este calculată prin procedeul mediei aritmetice, prin procedeul mediei mobile sau ajustînd seria după metoda celor mai mici pătrate.

Pentru demonstraţie se presupune că se dispune de următoarele date convenţionale, cu caracter de sezonalitate trimestrială, pentru anii 1975—1977.

Tabelul 10.10

Producţia trimestrială a întreprinderii *A* pe perioada 1975—1977

Trimestre	1975	1976	1977
I	20	30	40
II	70	90	110
III	150	180	240
IV	40	60	110
Total anual	280	360	500

10.6.1. Calculul sezonalităţii folosind metoda mediilor aritmetice

După cum se observă, an de an desfacerea acestui produs a înregistrat o creştere continuă. Dacă se analizează datele pe fiecare trimestru, se constată că există oscilaţii foarte mari, care însă, prin mecanismul formării mediilor, se compensează şi dispar atunci cînd se referă la mărimea acestor indicatori pe ani. Urmează, deci, că pentru a măsura sezonalitatea este necesar să se calculeze un raport între datele care sînt influenţate de valul sezonier şi media anuală care compensează abaterile într-un sens şi altul. Dar se poate observa că pe trimestre nu este aceeaşi variaţie între cei trei ani prezentaţi. Deci, şi aici va fi necesar să se determine, mai întîi, tendinţa pe fiecare trimestru, calculată ca o medie pe mai mulţi ani consecutivi şi

Scanned with OKEN Scanner

aceste medii parţiale trimestriale să se raporteze la media anuală calculată pe aceeaşi perioadă. Mediile trimestriale vor fi :

$$\bar{Y}_I = \frac{20 + 30 + 40}{3} = 30;$$

$$\bar{Y}_{II} = \frac{70 + 90 + 100}{3} = 90;$$

$$\bar{Y}_{III} = \frac{150 + 180 + 240}{3} = 190;$$

$$\bar{Y}_{IV} = \frac{40 + 60 + 110}{3} = 70.$$

Media anuală se poate calcula ca medie generală din cele patru medii parţiale :

$$\bar{Y}_0 = \frac{30 + 90 + 190 + 70}{4} = 95.$$

Pentru a măsura gradul de sezonalitate se face raportul dintre fiecare medie trimestrială (lunară) şi media generală :

$$I_s = \frac{\bar{Y}_i}{\bar{Y}_0} \cdot 100 \qquad (10.55)$$

Coeficienţii de sezonalitate trimestrială în exemplul luat vor fi :

$$I_{sI} = \frac{\bar{Y}_I}{\bar{Y}_0} \cdot 100 = \frac{30}{95} \cdot 100 = 31{,}6\%;$$

$$I_{sII} = \frac{\bar{Y}_{II}}{\bar{Y}_0} \cdot 100 = \frac{90}{95} \cdot 100 = 94{,}7\%;$$

$$I_{sIII} = \frac{\bar{Y}_{III}}{\bar{Y}_0} \cdot 100 = \frac{190}{95} \cdot 100 = 200\%;$$

$$I_{sIV} = \frac{\bar{Y}_{IV}}{\bar{Y}_0} \cdot 100 = \frac{70}{95} \cdot 100 = 74{,}7\%.$$

Deci, se poate aprecia că fenomenul în trimestrul II se apropie foarte mult de media anuală, cea mai mică desfacere realizîndu-se în trimestrul I, iar în trimestrul III rulajul este dublu faţă de media anuală. Pentru

Scanned with OKEN Scanner

a verifica calculul indicilor de sezonalitate, se face suma lor, care trebuie să fie egală cu numărul indicilor înmulţit cu o sută.

$$316 + 94{,}7 + 200 + 73{,}7 = 4 \cdot 100 = 400$$

Dacă sezonalitatea ar fi lunară, suma indicilor de sezonalitate ar fi egală cu 1 200.

Dar folosind acest procedeu s-a făcut abstracţie de faptul că termenii acestei serii dinamice sînt influenţaţi de legitatea lor internă de dezvoltare continuă de la o perioadă la alta. Această tendinţă de creştere continuă se poate urmări nu numai anual, ci şi pe trimestre corespunzătoare celor trei ani consecutivi. Or, prin calculul mediilor aritmetice trimestriale nu s-a ţinut seama şi de creşterile de volum de la o perioadă la alta, luîndu-se toţi termenii cu aceeaşi importanţă.

10.6.2. Calculul sezonalităţii folosind metoda mediilor mobile

Pentru a ţine seama şi de tendinţa obiectivă de dezvoltare a fenomenului, independent de valul sezonier, în teoria şi practica statistică se aplică, mai întîi, ajustarea termenilor pe baza mediilor mobile sau a ecuaţiilor de regresie.

În cazul mediilor mobile trebuie ţinut seama de periodicitatea termenilor, pentru a nu modifica tendinţa fenomenului. Cum periodicitatea termenilor este trimestrială, ajustarea seriei se va face pe baza unor medii mobile calculate din cîte patru termeni, iar pentru a fi centrate este necesar să se calculeze în două faze (medii mobile provizorii şi medii mobile definitive).

Mediile mobile provizorii:

$$\bar{y}_1 = \frac{20 + 70 + 150 + 41}{4} = 70; \qquad \bar{y}_5 = \frac{30 + 90 + 180 + 60}{4} = 90;$$

$$\bar{y}_2 = \frac{70 + 150 + 40 + 30}{4} = 72{,}5; \qquad \bar{y}_6 = \frac{90 + 180 + 60 + 40}{4} = 92{,}5;$$

$$\bar{y}_3 = \frac{150 + 40 + 30 + 90}{4} = 77{,}5; \qquad \bar{y}_7 = \frac{180 + 60 + 40 + 110}{4} = 97{,}5;$$

$$\bar{y}_4 = \frac{40 + 30 + 90 + 180}{4} = 85; \qquad \bar{y}_8 = \frac{60 + 40 + 110 + 240}{4} = 112{,}5;$$

$$\bar{y}_9 = \frac{40 + 110 + 240 + 110}{4} = 125.$$

Scanned with OKEN Scanner

După cum se poate observa, la calculul tuturor mediilor mobile sînt reprezentate toate trimestrele anului. Din această cauză, sezonalitatea a fost estompată şi valorile obţinute arată o creştere lină, continuă.

Dacă se plasează aceste medii pe o axă în care s-a înscris seria iniţială de date, atunci ele se vor plasa între termenii seriei, astfel :

20	70	150	41	30	90	180	60	40	110	240	110
y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
—	—	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	$\bar{y}_5$	$\bar{y}_6$	$\bar{y}_7$	$\bar{y}_8$	$\bar{y}_9$	
		70	72,5	77,5	85	90	92,5	97,5	112,5	125	

Deci, pe baza acestor medii mobile nu se poate face încă ajustarea şi atunci se calculează mediile mobile definitive :

$$\bar{\bar{y}}_1 = \frac{70 + 72{,}5}{2} = 71{,}25\,; \qquad \bar{\bar{y}}_5 = \frac{90 + 92{,}5}{2} = 91{,}25\,;$$

$$\bar{\bar{y}}_2 = \frac{72{,}5 + 77{,}5}{2} = 75\,; \qquad \bar{\bar{y}}_6 = \frac{92{,}5 + 97{,}5}{2} = 95\,;$$

$$\bar{\bar{y}}_3 = \frac{77{,}5 + 85}{2} = 81{,}25\,; \qquad \bar{\bar{y}}_7 = \frac{97{,}5 + 112{,}5}{2} = 105\,;$$

$$\bar{\bar{y}}_4 = \frac{85 + 90}{2} = 87{,}5\,; \qquad \bar{\bar{y}}_8 = \frac{112{,}5 + 125}{2} = 118{,}75.$$

Cu aceste medii mobile definitive se va ajusta seria empirică şi se va constata că variaţia sezonieră a fost înlăturată. Pentru a reveni la sezonalitate se va face raportul dintre termenii reali şi cei ajustaţi şi care vor folosi în continuare la determinarea indicelui de sezonalitate după acelaşi principiu ca şi la media aritmetică :

Tabelul 10.11

Calculul variaţiei sezoniere cu ajutorul mediilor mobile

Anii	Trimestre	Date iniţiale (y_i)	Date ajustate prin medii mobile ($\bar{\bar{y}}_i$)	Raportul între termenii reali şi cei ajustaţi $y_i/\bar{\bar{y}}_i$
A	B	1	2	3
1975	I	20	—	—
	II	70	—	—
	III	150	71,25	2,1
	IV	40	75	0,53
1976	I	30	81,25	0,35
	II	90	87,5	1,02
	III	180	91,25	1,97
	IV	60	95,0	0,63
1977	I	40	105,0	0,38
	II	110	118,75	0,92
	III	240	—	—
	IV	110	—	—

Scanned with OKEN Scanner

Pentru a urmări cu şi mai multă uşurinţă analogia de calcul al indicelui sezonalităţii cu procedeul folosit la mediile aritmetice se vor înscrie rezultatele într-un tabel asemănător, calculînd mediile parţiale trimestriale şi media generală anuală în care mediile trimestriale s-au obţinut ca medii aritmetice simple calculate pe cei trei ani consecutivi (vezi tabelul 10.12,

Tabelul 10.12

Calculul indicilor de sezonalitate prin ajustarea cu ajutorul mediilor mobile

Trimestrele	Anii			Mediile trimestriale ($\bar{Y}_t$)	Indici de sezonalitate $\frac{\bar{Y}_t}{\bar{Y}_0}$
	1975	1976	1977		
0	1	2	3	4	5
I	—	0,35	0,38	0,365	36,9
II	—	1,02	0,92	0,975	97,6
III	2,1	1,97	—	2,035	206,9
IV	0,53	0,63	—	0,58	58,6
	—	—	—	0,989	400,0
				$\bar{y}_0$	$\Sigma \frac{\bar{Y}_t}{\bar{Y}_0} \cdot 100$

col. 4), iar indicele de sezonalitate trimestrial, ca raport între media trimestrială şi cea anuală (vezi tabelul 10.12, col. 5).

Ajustarea seriei dinamice prin metoda mediilor mobile prezintă însă dezavantajul că nu se obţin termeni teoretici pentru toată seria, că faţă de numărul termenilor iniţiali prin ajustare se obţine, deci, o serie mai mică. În plus, folosind ajustarea pe baza mediilor mobile, nu se ţine seama de forma generală de dezvoltare. Dacă tendinţa de lungă durată din care se elimină variaţia sezonieră este liniară, atunci şi prin mediile mobile se va ţine seama de această tendinţă. Dacă, însă, tendinţa de lungă durată este curbilinie, atunci este de aşteptat că între coeficienţii determinaţi pe baza mediilor aritmetice şi a mediilor mobile şi cei obţinuţi şi prin ajustarea după ecuaţia unei curbe să existe abateri mari. Interpretarea lor nu se poate face decît dacă se reprezintă grafic seria şi se compară linia (curba) valorilor ajustate cu cea empirică.

10.6.3. Calculul sezonalităţii pe baza datelor ajustate prin metoda celor mai mici pătrate

Pentru a alege corect funcţia de ajustare se reprezintă grafic seria de date şi se compară pentru fiecare subperioadă a ciclului valorile în dinamică. Dacă se analizează variaţia trimestrială, atunci pe grafic se observă cum sînt dependente între ele valorile pentru fiecare trimestru, prezentate succesiv.

Scanned with OKEN Scanner

Reprezentînd grafic seria, se obţine (fig. 10.6) variaţia sezonieră.

Liniile care unesc valorile trimestriale ale celor trei ani arată că tendinţa de dezvoltare obţinută prin eliminarea influenţei sezoniere este de formă rectilinie.

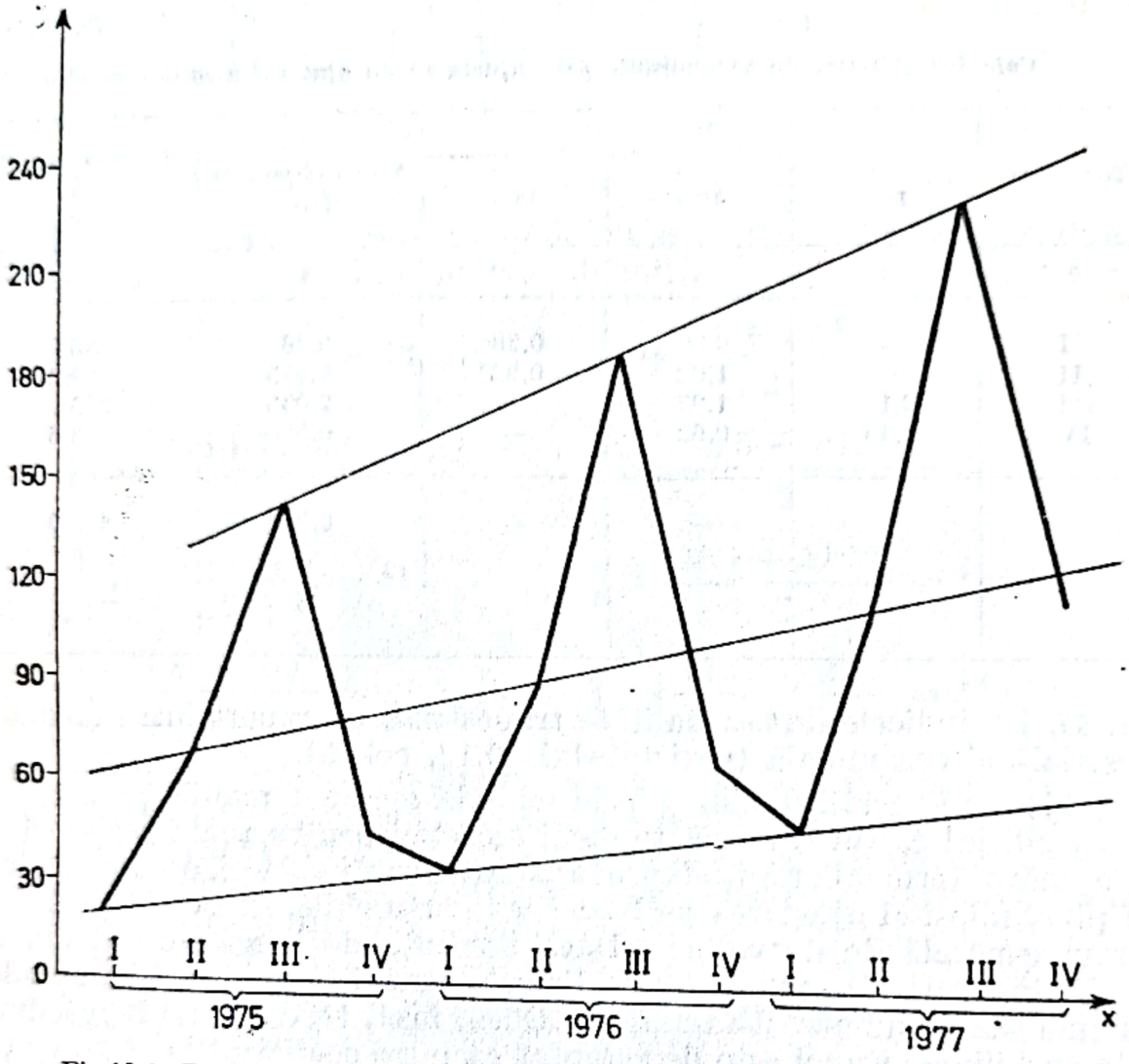


Fig.10.6. Producţia trimestrială a întreprinderii *A* pe perioada 1975—1977.

În continuare se va ajusta seria folosind ecuaţia liniei drepte ($Y_t = a + bt$) punînd condiţia ca suma variaţiei de timp să fie nulă ($\Sigma t_i = 0$). Se vor înscrie datele într-un tabel de calcul în care se va stabili şi raportul dintre termenii reali şi cei ajustaţi, pentru a putea calcula în continuare indicii de sezonalitate (vezi tabelul 10.13).

Pentru aflarea celor doi parametri a şi b se rezolvă sistemul:

$$\begin{aligned} 12\ a &= 1\,140 \\ 572\ b &= 2\,420 \end{aligned}$$

de unde:

$$a = \frac{1\,140}{12} = 95$$

$$b = \frac{2\,420}{572} = 4{,}23.$$

Scanned with OKEN Scanner

Tabelul 10.13

Calculul variației sezoniere pe baza datelor ajustate prin ecuația liniei drepte

Datele inițiale (y_i)	t	t^2	ty	$Y_t = a + bt$ $Y_t = 95 + 4{,}23\,t$	$\frac{y_i}{Y_t}$
1	2	3	4	5	6
20	−11	121	− 220	48,47	0,41
70	− 9	81	− 630	56,93	1,22
150	− 7	49	−1 050	65,39	2,29
40	− 5	25	− 200	73,85	0,54
30	− 3	9	− 90	82,31	0,36
90	− 1	1	− 90	90,77	0,99
180	+ 1	1	+ 180	99,23	1,81
60	+ 3	9	+ 180	107,69	0,55
40	+ 5	25	+ 200	116,15	0,34
110	+ 7	49	+ 770	124,61	0,89
240	+ 9	81	+2 160	133,07	1,80
110	+11	121	+1 210	143,47	0,77
1 140	$\Sigma t = 0$	572	2 420	1 141,84	

Prin ajustare (vezi coloana 5, tabelul 10.13) — ca și în cazul mediilor mobile — se poate observa că variația sezonieră a fost înlăturată și s-a obținut o creștere lină de la un trimestru la altul. Pentru a reveni la sezonalitate și în acest caz se face raportul dintre termenii reali (care exprimă influența tuturor factorilor inclusiv a valului sezonier) și termenii ajustați (care se obțin prin considerarea tuturor factorilor ca fiind cu acțiune constantă pe toată perioada manifestată sub forma unei tendințe liniare de dezvoltare de la un trimestru la altul).

Pentru a măsura gradul de sezonalitate cu luarea în considerare a tendinței liniare, se vor înscrie și aceste raporturi într-un tabel în care se vor calcula mediile trimestriale și indicii de sezonalitate (vezi tabelul 10.14).

Tabelul 10.14

Calculul indicilor de sezonalitate prin ajustarea liniară

Trimestre	1975	1976	1977	Media trimestrială ($\bar{y}_i$)	Indicii de sezonalitate $\frac{\bar{y}_i}{\bar{y}_0}$
0	1	2	3	4	5
I	0,41	0,36	0,34	0,37	0,38
II	1,22	0,99	0,89	1,00	1,01
III	2,29	1,81	1,80	1,97	1,97
IV	0,54	0,55	0,77	0,63	0,64
	—	—	—	0,99	4,00

Scanned with OKEN Scanner

Folosind trei procedee s-au obţinut trei serii de indici de sezonalitate. Pentru a putea compara rezultatele obţinute se vor înscrie într-un tabel cele trei serii de indici ai sezonalităţii :

Tabelul 10.15

Indicii sezonalităţii trimestriale pe perioada 1975—1977

Trimestrul	Indicii sezonalităţii calculaţi :		
	din termenii neajustaţi	din termenii ajustaţi	
		prin medii mobile	prin ecuaţia liniei drepte
0	1	2	3
I	31,6	36,9	38,0
II	94,7	97,6	101,0
III	200,0	206,9	197,0
IV	73,7	58,6	64,0
Total	400,0	400,00	400,0

Din tabelul 10.15, ca şi din grafic (fig. 10.7) rezultă că dacă se ţine seama şi de tendinţa de dezvoltare a fenomenului, deci se ajustează în prealabil seria de date pentru a înlătura influenţa celorlalţi factori şi după aceea se calculează gradul de sezonalitate, atunci între indicii de sezonalitate ai trimestrelor I şi IV abaterea este mai mică, accentuîndu-se şi mai mult sezonalitatea către trimestrele III şi II. Aceasta confirmă faptul că tendinţa de creştere liniară nu s-a manifestat uniform pe fiecare trimestru.

Mai sugestiv se compară cu ajutorul reprezentării grafice.

Gradul de sezonalitate se prezintă grafic printr-o diagramă polară (fig. 10.7).

10.7. FOLOSIREA DATELOR STATISTICE PENTRU FUNDAMENTAREA CALCULELOR DE PROGNOZĂ

În teoria şi practica statistică, în mod frecvent se pune problema de a găsi, pe baza interpretării unor serii de date statistice, care va fi nivelul unui fenomen într-o etapă următoare.

Problema extrapolării seriilor de date statistice în domeniul fenomenelor sociale şi economice este deosebit de dificilă. Este cunoscut faptul că fenomenele sociale şi economice sînt deosebit de dinamice, că asupra lor influenţează un număr mare de cauze, a căror influenţă poate fi foarte variată în raport cu condiţiile în care ele se produc. Dezvoltarea continuă a forţelor de producţie, în interdependenţă strînsă cu relaţiile de producţie, are ca efect modificări substanţiale cantitative şi calitative în cadrul fenomenelor şi proceselor sociale. Aceasta înseamnă că extrapolarea datelor statistice nu poate fi redusă numai la o aplicare strictă a unui procedeu sau altuia de calcul statistic, ci trebuie să se ţină seama de condiţiile

Scanned with OKEN Scanner

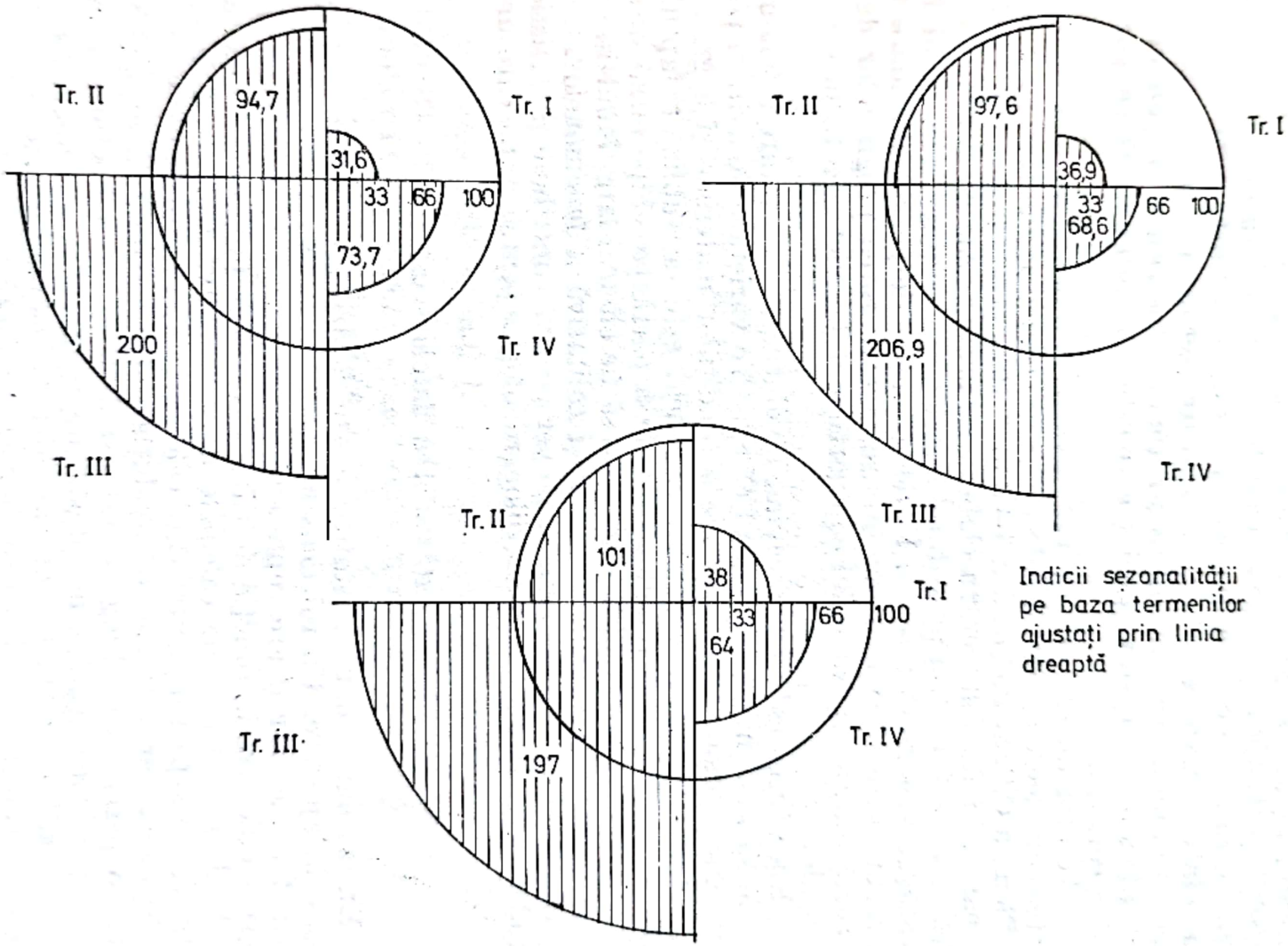


Fig. 10.7. Variația sezonieră a producției întreprinderii A în perioada 1975 – 1977.

concrete în care urmează să se producă fenomenele respective, cu alte cuvinte, să se efectueze o analiză teoretică care precede, de regulă, orice calcul statistic.

În socialism, statistica de stat trebuie să furnizeze datele necesare fundamentării planurilor de perspectivă și defalcării acestora în planuri curente. Pentru aceasta este necesar să se analizeze seriile statistice de date pe perioada expirată, să se determine tendința obiectivă de dezvoltare a acestora și să se aprecieze în ce măsură această tendință se va păstra și în viitor. De regulă, în extrapolarea seriilor de date statistice se stabilesc mai multe variante și, în funcție de interpretarea economică și politică a fenomenelor, se va decide care este varianta optimă ce corespunde noilor condiții.

Atenția de care se bucură această problemă în țara noastră, din partea conducerii de partid și de stat, este confirmată de Hotărîrea adoptată de Plenara C.C. al P.C.R. din 5—6 mai 1971 privind constituirea Comisiei de partid și de stat pentru elaborarea prognozelor de dezvoltare economică și socială, a cărei președinte a fost ales secretarul general al P.C.R., tovarășul Nicolae Ceaușescu. În documentele de partid ulterioare au fost prevăzute sarcini exprese cu privire la elaborarea prognozelor de lungă și scurtă durată în domeniul compartimentelor vieții economice și social-culturale.

Elaborarea calculelor privind nivelurile de dezvoltare a fenomenelor trebuie să aibă în vedere faptul că există o varietate de „viitori posibili" ca urmare a modificării continue a condițiilor materiale și a gradului de interacțiune dintre factori. De asemenea, este de subliniat faptul că la elaborarea calculelor de prognoză trebuie să participe echipe mixte de specialiști pentru care orizontul de prognoză să fie cît mai larg, luînd în considerare întreaga modificare cantitativă și calitativă a fenomenelor.

La fundamentarea calculelor de perspectivă contribuie și statistica prin metodele sale de extrapolare a seriilor cronologice pentru etapa care urmează.

10.7.1. Extrapolarea datelor statistice

Extrapolarea datelor statistice are la bază metodele și procedeele folosite la ajustare. Un rol deosebit în interpretarea tendinței de dezvoltare a fenomenelor îl are reprezentarea grafică, care constituie unul din criteriile obiective în funcție de care se va alege procedeul de extrapolare.

Analiza care precede calculele de perspectivă trebuie să ia în considerare unele aspecte, printre care cele mai importante sînt :

— cunoașterea formei de evoluție a fenomenului în etapa expirată, pentru o perioadă de lungă durată și divizată pe subperioade, dacă din analiza statistică rezultă modificări de tendință;

— identificarea factorilor care determină evoluția fenomenelor studiate și măsura în care influența lor se păstrează sau se modifică în etapa ce urmează;

— alegerea perioadei la care se referă seria cronologică analizată, astfel încît să fie semnificativă pentru dezvoltarea fenomenului și să cuprindă un număr suficient de termeni pentru a putea face o interpretare corectă.

Scanned with OKEN Scanner

Numărul variantelor calculate pe baza extrapolării seriilor cronologice este determinat, în principal, de forma de dezvoltare preconizată, de mărimea parametrilor folosiţi şi de modul în care este considerat fenomenul în sistemul de prognoză.

În raport cu forma de dezvoltare preconizată se pot obţine atîtea variante cîte procedee s-au folosit la ajustarea seriei pe perioada expirată.

În raport cu mărimea parametrilor folosiţi se pot obţine mai multe variante în care se păstrează aceleaşi relaţii cantitative între variabila pentru care se fac calcule de perspectivă şi variabila de timp ; deci, se folosesc aceleaşi valori ale parametrilor ca şi în perioada expirată, şi o serie de variante cu valori modificate ale parametrilor, în funcţie de modificarea cantitativă a condiţiilor de bază în etapa următoare.

În raport cu modul în care este privit fenomenul în sistemul de prognoză, se poate obţine o serie de variante în care fenomenul este studiat ca o variabilă independentă în timp de factorii săi social-economici, precum şi o serie de variante în care variabila este dependentă, în timp, de unul sau mai mulţi factori social-economici.

Aceste consideraţii arată că termenii obţinuţi din extrapolarea unei serii statistice trebuie apreciaţi ca termeni probabili, care pot prezenta erori faţă de valorile reale ce se vor obţine în viitor. Aceste erori sînt, în principal, de două feluri : erori obiective de extrapolare şi erori provenite din modelul ales. Erorile obiective de extrapolare apar ca urmare a faptului că şi în viitor factorii accidentali îşi pot modifica gradul lor de influenţă faţă de linia (curba) de tendinţă ce exprimă acţiunea sistematică a factorilor esenţiali. Erorile de extrapolare, provenite din alegerea modelului, apar ca urmare a faptului că în viitor pot să apară modificări cantitative şi calitative ce nu pot fi previzibile cu certitudine în momentul de calcul şi care pot modifica radical tendinţa generală specifică perioadei anterioare.

Statistica are, deci, rolul de a furniza mai multe variante ce rezultă din extrapolarea seriilor cronologice. Decizia urmează să se ia pe baza analizei politice şi economice de către organele de conducere.

10.7.1.1. EXTRAPOLAREA PRIN PROCEDEE MECANICE

În calculele statistice de perspectivă se utilizează destul de frecvent extrapolarea datelor statistice pe baza sporului mediu şi a ritmului mediu de creştere.

La extrapolarea pe baza sporului mediu şi a ritmului mediu de creştere se porneşte de la ipoteza că s-ar păstra aceeaşi bază de calcul şi că fenomenele ar evolua în aceleaşi condiţii ca şi în perioada expirată, păstrînd aceeaşi tendinţă de apropiere către sporurile cu baza în lanţ (pentru fenomenele care au tendinţa de creştere de forma unei progresii aritmetice) şi de apropiere către indicii de creştere cu bază în lanţ (cînd tendinţa de dezvoltare este către forma unei progresii geometrice).

Scanned with OKEN Scanner

Pentru alegerea formei de evoluţie, datele se reprezintă grafic, ca şi în cazul ajustării ; se face ajustarea pe perioada expirată şi aplicînd în continuare sporul mediu (respectiv indicele mediu) se determină valorile seriei extrapolate pentru perioada care interesează.

Pentru a face distincţie între termenii ajustaţi (Y_t) şi cei extrapolaţi — care sînt consideraţi tot termeni teoretici — se vor nota termenii extrapolaţi cu Y', iar variabila de timp cu t'. Variabila de timp t' formată din valorile de timp ale seriei empirice : $t_0, \ldots, t_n$, prelungită cu valorile de timp r, adică ale seriei extrapolate $t_n \ldots t_r$.

Deci, formulele de calcul vor fi :

— pentru extrapolarea pe baza sporului mediu :

$$Y'_t = y_0 + t'\overline{\Delta}; \tag{10.56}$$

— pentru extrapolarea pe baza indicelui mediu de creştere :

$$Y'_t = y_0 \cdot \bar{I}'^{t'} \tag{10.57}$$

Aceste formule se aplică atunci cînd se folosesc valorile parametrilor ($\overline{\Delta}$, $\bar{I}$) din perioada expirată. În cazul cînd se modifică sporul mediu sau indicele mediu de creştere, formulele se modifică cu un coeficient K, astfel :

$$Y'_t = y_0 + t'\overline{\Delta}', \tag{10.58}$$

în care :

$$\overline{\Delta}' = K\overline{\Delta} \tag{10.59}$$

$$Y'_t = y_0 \bar{I}'^{t'} \tag{10.60}$$

în care există relaţia

$$\bar{I}' = K\bar{I}. \tag{10.61}$$

Coeficientul K poate să fie mai mare sau mai mic decît 1.

Dacă $K < 1$, atunci înseamnă că se reduce variaţia medie absolută sau relativă, după cum se aplică la primul sau la al doilea procedeu.

Dacă $K > 1$, atunci înseamnă că valoarea parametrilor folosiţi în extrapolare este mai mare decît în perioada de analiză.

De cele mai multe ori, pentru fenomenele care au tendinţa de creştere continuă se aplică un coeficient $K < 1$, în cazul cînd se foloseşte extrapolarea de forma progresiei geometrice (de exemplu, indicele mediu anual de creştere) şi un coeficient $K > 1$, cînd se apreciază că fenomenul va urma unei progresii aritmetice (de exemplu, pentru sporul mediu).

În continuare, folosind aceleaşi date ca şi la ajustare şi aceleaşi valori pentru cei doi parametri : sporul mediu şi indicele mediu de creştere se vor obţine valorile extrapolate pentru perioada 1976—1980 (vezi tabelul 10.16).

Scanned with OKEN Scanner

Se poate observa că s-au obținut valori foarte diferite între ele. Pentru a putea compara noile date obținute se va calcula creșterea totală ce se va obține folosind cele două variante de calcul, comparînd valoarea extrapolată pentru anul 1980 cu cea pentru anul 1976.

Pe baza sporului mediu, creșterea totală calculată în perspectivă va fi :

$$Y'_{1980} - Y_{1976} = 71\,974{,}1 - 57\,371{,}6 = 14\,597{,}50 \text{ mil.Kwh.}$$

Pe baza indicelui mediu de creștere, sporul total va fi :

$$Y'_{1980} - Y_{1976} = 94\,886{,}3 - 60\,194{,}4 = 34\,691{,}90 \text{ mil.Kwh.}$$

Tabelul 10.16

Estimarea valorilor producției de energie electrică a R.S. România pe perioada 1976—1980, în funcție de sporul mediu și indicele de creștere al perioadei 1965—1975

Anii	Variația de timp	Date extrapolate (mil. Kwh):	
		prin sporul mediu $Y' = 17\,215 + 3\,650{,}6\ t'_i$	prin indicele mediu $y' = 17\,215 \cdot 1{,}1205\ t_i$
A	1	2	3
1976	11	$17\,215+(11 \cdot 3650{,}6) = 57\,371{,}6$	$17\,215 \cdot 1{,}1205^{11} = 60\,194{,}4$
1977	12	$17\,215+(12 \cdot 3650{,}6) = 61\,022{,}2$	$17\,215 \cdot 1{,}1205^{12} = 67\,447{,}8$
1978	13	$17\,215+(13 \cdot 3650{,}6) = 64\,672{,}2$	$17\,215 \cdot 1\,1205^{13} = 75\,575{,}3$
1979	14	$17\,215+(14 \cdot 3650{,}6) = 68\,323{,}4$	$17\,215 \cdot 1{,}1205^{14} = 84\,682{,}1$
1980	15	$17\,215+(15 \cdot 3650{,}6) = 71\,974{,}1$	$17\,215 \cdot 1{,}1205^{15} = 94\,886{,}3$

Deci, dacă producția de energie electrică va crește de forma unei progresii geometrice, sporul pe perioada 1976—1980 va fi de 2,28 ori $\left(\frac{34{,}691{,}9}{14\,597{,}5}\right)$ mai mare față de creșterea după o progresie aritmetică folosind valorile celor doi parametri din perioada 1965—1975.

10.7.1.2. EXTRAPOLAREA PE BAZA METODELOR ANALITICE DE CALCUL

Și în acest caz se consideră că s-ar păstra aceeași tendință și aceleași condiții și în etapa pentru care se fac calcule de perspectivă.

Pentru extrapolarea pe baza metodelor analitice de calcul se pune, în primul rînd, condiția ca datele să se determine astfel încît să nu modifice originea variației de timp care este în mijlocul seriei cronologice și pentru care $\Sigma t_i = 0$. Deci, variația de timp se extinde în ambele sensuri, deși interesează numai tendința obținută prin extinderea seriei pentru perioada următoare. Aceasta înseamnă că, folosind metodele de calcul, bazate pe metoda celor mai mici pătrate, nu se poate separa extrapolarea

Scanned with OKEN Scanner

de ajustare. Se face ajustarea datelor pe perioade expirată și se calculează numai valorile pentru etapa care urmează.

Deoarece, la seria privitoare la producția de energie electrică s-au folosit trei funcții de ajustare, se vor calcula datele extrapolate după funcția liniară, funcția exponențială și parabola crescătoare de gradul 2, cu următoarele ecuații medii de estimare:

— pentru funcția liniei drepte:

$$\bar{Y}_{1965-1975} = 35\,424{,}6 + 3\,643\,t_i;$$

— pentru funcția exponențială:

$$\bar{Y}_{1965-1975} = 33\,400 \cdot 1{,}118^{t_i};$$

— pentru funcția parabolică de gradul 2:

$$\bar{Y}_{1965-1975} = 35\,488 + 3\,643{,}2\,t_i - 6{,}34\,t_i^2.$$

Și în acest caz se vor folosi aceleași notații, adică termenii extrapolați se vor nota cu Y'_t, astfel:

$$Y'_t = a + bt' \tag{10.62}$$

$$Y'_t = ab^{t'} \tag{10.63}$$

$$Y'_t = a + bt' + ct'^2. \tag{10.64}$$

Pentru perioada 1976—1980, prin extrapolarea pe baza metodelor analitice se vor obține datele din tabelul 10.17.

Comparînd cele trei șiruri de valori extrapolate, folosind cele trei funcții, se poate aprecia că valorile celui de-al treilea șir se apropie de acelea din primul șir ca urmare a faptului că, în ecuația parabolei de gradul 2, cel de-al treilea parametru este cu valoare negativă. Comparînd datele extrapolate pe baza procedeelor mecanice (sporul mediu, indicele mediu) se observă că abateri mici apar între valorile calculate pe baza sporului mediu și cele calculate pe baza funcției liniei drepte, pe de o parte, și abateri mici între valorile calculate pe baza indicelui mediu și a ecuației de tendință exponențială, pe de altă parte. Reiese de aici, că, în practică, dacă nu se dispune de toată seria de date, ci numai de valorile înregistrate la diferite perioade de timp, se poate folosi extrapolarea pe baza sporului mediu și a indicelui mediu, care nu sînt legate de cunoașterea valorilor întregii serii și de ajustarea prealabilă a acestora.

Scanned with OKEN Scanner

Tabelul 10.17

Estimarea valorilor producției de energie electrică a R.S. România pe perioada 1976—1980, prin extrapolarea parametrilor funcției liniare, exponențiale și parabolice de gradul 2 din perioada 1965—1975

Anii	Variația de timp	Date extrapolate: $Y'_t = 35\,424{,}6 + 3\,643\,t_i$	$Y'_t = 33\,400 \cdot 1{,}118^{t_i}$	$Y'_t = 35\,488 + 3\,643{,}2\,t_i - 6{,}34\,t^2_i$
A	1	2	3	4
1976	6	57 282,6	60 060,3	57 118,8
1977	7	60 952,6	67 147,0	60 679,7
1978	8	64 568,6	75 070,5	64 227,8
1979	9	68 211,6	83 928,9	67 763,3
1980	10	71 854,6	93 832,5	71 286,0

Avînd în vedere faptul că perioada 1976—1980 formează anii unui cincinal, este necesar să se calculeze care ar fi creșterile totale absolute și relative ale producției de energie electrică, precum și ritmurile de creștere medie anuală dacă s-ar adopta una din variante (vezi tabelul 10.18).

Rezultă că, pentru a imprima producției de energie electrică pe perioada 1976—1980 o creștere exponențială, ar trebui ca ritmul mediu anual de creștere să fie în jur de 12%, asigurîndu-se astfel o creștere de circa 1,5 ori a producției prevăzute față de primul an al cincinalului. Dacă se imprimă producției de energie electrică o creștere liniară, atunci ritmul mediu de creștere va fi circa 6% realizîndu-se o creștere totală de 125% a producției anului 1980 față de producția calculată pentru anul 1976.

Tabelul 10.18

Modificarea ratei absolute și a celei relative, a creșterii producției de energie electrică pe perioada 1976—1980, în funcție de procedeul folosit la extrapolare

Procedeul folosit la extrapolare	Sporul total 1980/1976 Δ	Sporul mediu 1976—1980 $\overline{\Delta}$	Indicele cu bază fixă 1980/1976 I	Indicele mediu 1976—1980 $\overline{I}$
A	1	2	3	4
— sporul mediu	14 597,50	3 650,6	125,5 %	103,8 %
— indicele mediu	34 691,90	8 673,0	157,6 %	112,05 %
— ecuația dreptei	14 572,0	3 643,0	125,44 %	105,83 %
— ecuația funcției exponențiale	33 792,5	8 448,1	156,23 %	111,80 %
— parabola de gradul 2	14 167,2	3 541,8	124,80 %	105,7 %

Pentru verificarea calculelor de prognoză statistică obținute prin extrapolare se va calcula eroarea efectivă ca abatere între valoarea estimată și mărimea producției obținută efectiv în anii 1976 și 1977.

Din datele Anuarului Statistic pe 1979 (p. 177), producția de energie electrică a anului 1976 a fost de 58 266 mil. Kwh, în anul 1977 a fost de 59 858 mil. Kwh, iar în 1978 de 64 255 mil. Kwh, ceea ce înseamnă că

Scanned with OKEN Scanner

pentru primii trei ani ai cincinalului creşterea producţiei se apropie de forma unei creşteri liniare, faţă de care se obţin cele mai mici abateri.

Modelele aditive şi multiplicative prezentate pînă aici au considerat producţia ca o variabilă dependentă numai de timp. În teoria şi practica statistică se pot folosi şi modele bazate pe corelaţia indicatorilor în dinamică.

Dacă fenomenele se analizează în interdependenţă cu variaţia factorilor care le determină în dinamică, atunci se aleg caracteristicile factoriale care sînt previzibile numai în funcţie de timp şi apoi, apreciind că s-ar păstra aceeaşi tendinţă de realizare a legăturii, să se calculeze valorile de perspectivă şi pentru caracteristica dependentă. De exemplu, pentru o caracteristică rezultativă, analizată în funcţie de doi factori cu care este legată liniar, se calculează, mai întîi, ecuaţia medie de tendinţă pe perioada expirată, adică :

$$Y = a_0 + a_1 x_1 + a_2 x_2,$$

în care :

x_1, x_2 sînt variabile factoriale luate în timp.

Separat se analizează tendinţa de dezvoltare a fiecărei variabile independente şi se extrapolează cele două serii :

$$X_1' = a_{x_1} + b_{x_1} t' ; \tag{10.65}$$

$$X_2' = a_{x_2} + b_{x_2} t' \tag{10.66}$$

Aceste valori se introduc în ecuaţia funcţiei de regresie multiplă, calculîndu-se valorile extrapolate ale variabilei y, după valorile modificate ale celor doi factori :

$$Y'_{x_1 x_2} = a + a_1 x_1' + a_2 x_2' \tag{10.67}$$

Au fost prezentate cîteva modele care pot fi folosite în calculele previzionale. Fără îndoială că pentru a lua decizia corespunzătoare se vor analiza condiţiile de producţie, respectiv capacitatea de producţie actuală şi cea care urmează să fie dată în folosinţă, nivelul de productivitate a muncii etc., care vor putea permite alegerea variantei ce corespunde cel mai bine noilor condiţii de producţie.

10.7.2. Interpolarea seriilor cronologice

În mod frecvent, în practica statistică se poate pune problema de a găsi una sau mai multe valori care lipsesc în interiorul unei serii statistice. În acest caz este necesar să se recurgă la operaţia de interpolare. Pentru interpolarea statistică se pot folosi mai multe procedee, printre care cele mai frecvente se bazează pe relaţiile dintre parametrii care definesc seria : sporul mediu, indicele mediu şi valorile cunoscute ale seriei.

Operaţia de interpolare poate fi asimilată celei de ajustare cu deosebirea că, în acest caz, nu se cunosc valorile empirice pentru perioada în

Scanned with OKEN Scanner

care se analizează dinamica fenomenelor şi, în consecinţă, nu se poate determina forma de evoluţie specifică perioadei respective. Cunoscîndu-se doar două valori între două limite date, înseamnă că între ele se poate admite o relaţie de aditivitate şi atunci se recurge la forma unei progresii aritmetice sau o relaţie de multiplicativitate şi se recurge la forma unei progresii geometrice.

Necesitatea de a interpola o serie de date poate să apară şi în cazul unor calcule de prognoză.

Interpolarea seriilor cronologice se foloseşte, de obicei, în cazul prognozelor normative, cînd se stabileşte nivelul la care trebuie să ajungă un indicator de plan într-o anumită perioadă de timp şi trebuie să fie calculate valorile estimate pentru fiecare unitate intermediară de timp.

De exemplu, la Congresul al XI-lea al P.C.R. s-a stabilit ca producţia de ciment a R.S. România să ajungă în anul 1990 la 25—28 mil. tone. Din datele publicate în Anuarul statistic pe anul 1977, rezultă că în anul 1975 în România se produceau 11 525 mii tone. Se pune problema de a se stabili valorile estimate ale producţiei anuale de ciment din ţara noastră, care să asigure atingerea nivelului prognozat al producţiei anului 1990, încît să se asigure o creştere treptată a producţiei anuale.

Pentru aceasta este necesar să se calculeze care trebuie să fie sporul mediu şi ritmul mediu anual al producţiei de ciment pe perioada 1975—1990 spre a se asigura atingerea nivelului prognozat.

Pentru limita minimă, $Y_{1990} = 25\,000$ mii tone,

$$\overline{\Delta}_{1975-1990} = \frac{25\,000 - 11\,525}{15} = 898{,}3 \text{ mii tone/anual};$$

$$\bar{I}_{1975-1990} = \sqrt[15]{\frac{25\,000}{11\,525}} = \sqrt[15]{2{,}169}$$

$$\log \bar{I}_{1975-1990} = \frac{\log 2{,}169}{15} = \frac{0{,}33626}{15} = 0{,}02242,$$

de unde:

$$\bar{I}_{1975-1990} = 1{,}053 \text{ sau } 105{,}3\,\%$$

Pentru limita maximă, $Y_{1990} = 28\,000$ mii tone,

$$\overline{\Delta}_{1975-1990} = \frac{28\,000 - 11\,525}{15} = 1\,098 \text{ mii tone/anual},$$

$$\bar{I}_{1975-1990} = \sqrt[15]{\frac{28\,000}{11\,525}} = \sqrt[15]{2{,}429}$$

$$\log \bar{I}_{1975-1990} = \frac{\log 2{,}429}{15} = \frac{0{,}38543}{15} = 0{,}02570,$$

Scanned with OKEN Scanner

de unde:

$$\bar{I}_{1975-1990} = 1{,}061 \text{ sau } 106{,}1\%.$$

Valorile interpolate ale seriei se pot calcula considerîndu-se ca fiind termenii unei progresii aritmetice cu raţia egală cu sporul mediu sau a unei progresii geometrice cu raţia egală cu indicele mediu de creştere (vezi tabelul 10.19).

Valorile înscrise în tabelul 10.19 s-au calculat ca şi în cazul extrapolării prin procedee mecanice, cu deosebirea că aici s-au estimat sporul mediu şi indicele mediu care să asigure nivelul stabilit pentru producţia de ciment a anului 1990. La producţia cunoscută a anului 1975 s-a adăugat succesiv sporul mediu care a fost estimat admiţînd o creştere sub formă de progresie aritmetică, adică cu o sarcină absolută distribuită egal pe fiecare an, sau s-a multiplicat succesiv cu indicele mediu admiţîndu-se astfel o ipoteză de creştere sub forma unei progresii geometrice, ceea ce ar coresTabelul 10.19

Valorile estimate ale producţiei de ciment pe perioada 1975—1980, pe baza datelor stabilite la Congresul al XI-lea, folosind sporul mediu şi indicele mediu de creştere

Anii	Pentru o producţie de ciment a anului 1990			
	25 000 mii tone		28 000 mii tone	
	asigurînd un spor mediu de 898,3 mii	asigurînd un indice mediu = 1,053	asigurînd un spor mediu = 1 098 mii	asigurînd un indice mediu = 1,061
A	1	2	3	4
1975	11 525	11 525	11 525	11 525
1976	12 423	12 135,8	12 623	12 228,0
1977	13 321	12 779	13 721	12 973,9
1978	14 219	13 477,4	14 819	13 765,3
1979	15 117	14 191,7	15 917	14 605,0
1980	16 015	14 943,9	17 015	15 495,9
1981	16 913	15 735,9	18 113	16 441,2
1982	17 811	16 569,9	19 211	17 444,1
1983	18 709	17 448,1	20 309	18 508,2
1984	19 607	18 372,9	21 407	19 637,2
1985	20 505	19 346,7	22 505	20 835,1
1986	21 403	20 372,1	23 603	22 106,0
1987	22 301	21 451,8	24 704	23 454,5
1988	23 199	22 588,8	25 799	24 885,2
1989	24 097	23 786,0	26 897	26 403,2
1990	24 995	25 046,7	27 995	28 013,8

punde unor creşteri absolute din ce în ce mai mari. Pe măsura trecerii timpului pentru care s-a elaborat prognoza se vor calcula erorile efective ca abateri între valorile reale şi cele prognozate după ambele variante şi se vor lua măsuri pe parcurs de schimbare fie a ritmului mediu, fie a sporului mediu anual pentru a asigura îndeplinirea planului propus pentru anul 1990.

Scanned with OKEN Scanner

10.7.3. Calcule de perspectivă pe plan teritorial

În teoria și practica statistică, privitor la comparațiile pe plan teritorial, se pun destul de des probleme legate de interpretarea unor ritmuri diferite de dezvoltare, pe baza cărora se stabilesc relații între județe și relații între județe și întreaga țară. Astfel se pot pune probleme legate de determinarea gradului de devansare, de timpul în care nivelul unui indicator se poate dubla sau tripla, de perioada de timp după care un județ va ajunge din urmă un altul din punct de vedere absolut, sau ce ritm de dezvoltare este necesar pentru a ajunge la un anumit nivel, după expirarea unei anumite perioade de timp.

Indicii de devansare se calculează fie pentru două unități teritoriale, fie în interiorul aceleiași colectivități, pentru două variabile prezentate în timp.

Spre exemplu, se poate pune problema de a stabili în cadrul unui județ — de pildă Ialomița — care este indicele de devansare a dinamicii producției agricole de către dinamica producției industriale pe perioada 1965—1975. Pentru aceasta se vor compara indicii de creștere, efectuîndu-se un raport între indicii de creștere cu bază fixă a celor două variabile. Pentru a se face comparația, se notează cu y variabila care are o dinamică mai rapidă și cu z variabila care are o dinamică mai lentă.

$$I^{y}_{n/0} = \frac{y_n}{y_0} \text{ — indicele de raportat;} \qquad (10.68)$$

$$I^{z}_{n/0} = \frac{z_n}{z_0} \text{ — indicele bază de comparare,} \qquad (10.69)$$

$$I_{y/z} = \frac{I^{y}_{n/0}}{I^{z}_{n/0}} \text{ — indicele de devansare} \qquad (10.70)$$

După datele publicate în Anuarul statistic al R.S. România pe anul 1976 (pag. 94) producția globală industrială a fost de 5,01 ori mai mare în 1975 față de 1965, iar producția globală agricolă a crescut numai de 2,09 ori.

Indicele de devansare a dinamicii producției globale agricole de către dinamica producției industriale va fi:

$$I_{y/z} = \frac{5{,}01}{2{,}09} = 2{,}432 \text{ ori sau } 243{,}2\%$$

Aceasta înseamnă că, pe perioada 1965—1975, în județul Ialomița ritmul de creștere al producției industriale a devansat cu 143,2% ritmul de creștere al producției globale agricole. De remarcat este și faptul că pentru cei doi indicatori — producția globală industrială și producția globală agricolă — județul Ialomița se plasează diferit față de valoarea indicilor pe total.

Scanned with OKEN Scanner

Pe întreaga țară indicele dinamicii producției globale industriale este de 322, iar cel al producției globale agricole este 293. Deci, se poate calcula un indice de devansare numai pentru dinamica producției globale industriale a județului Ialomița față de dinamica pe întreaga țară.

Dacă județul Ialomița se notează cu A și întreaga țară cu T, atunci indicele de devansare a ritmului de creștere al producției industriale a țării de către ritmul de creștere al producției globale industriale din județul Ialomița, va fi :

$$I_{A/T} = \frac{I^A_{n/o}}{I^T_{n/o}} = \frac{5{,}01}{3{,}22} = 1{,}56.$$

Deci, dinamica producției globale industriale a județului Ialomița pe perioada 1965—1975 a devansat dinamica aceluiași indicator cu 56% pe întreaga țară și cu 143,2 ori dinamica producției globale agricole din același județ.

O altă problemă legată de extrapolarea seriilor dinamice este aceea în care, cunoscîndu-se ritmul mediu actual de dezvoltare a unui județ, sau al țării, se cere să se calculeze *după ce perioadă de timp fenomenul s-ar modifica de un număr de ori sau de cîte ori se va modifica fenomenul după un timp dat*, în ambele cazuri păstrîndu-se aceleași condiții de dezvoltare.

Presupunînd dat indicele mediu de creștere, înseamnă că se admite o relație de progresie geometrică între termenii seriei, care se poate scrie astfel :

$$\bar{I}^x = k, \qquad (10.71)$$

în care :

$\bar{I}$ reprezintă indicele mediu de creștere cu care se estimează variația în dinamică a fenomenelor ;

x — numărul de ani după care se va produce schimbarea de k ori ;

k — coeficientul de schimbare a fenomenului după trecerea celor x ani.

Din exemplul prezentat anterior se știe că indicele mediu de creștere a producției de energie electrică pe perioada 1965—1975 a fost de 1,121 ori. Se poate pune problema ca, păstrînd același indice mediu de dezvoltare, să se determine după ce număr de ani producția țării se va dubla sau tripla în raport cu valoarea producției din momentul de calcul (1975) :

— pentru dublare :

$$1{,}121^x = 2$$

deci : $x \log 1{,}121 = \log 2$,

de unde :

$$x = \frac{0{,}30103}{0{,}04961} = 6{,}1 \text{ ani} ;$$

— pentru triplare :

$$1{,}121^x = 3$$

Scanned with OKEN Scanner

deci : $x \log 1{,}121 = \log 3$

$$0{,}04961\ x = 0{,}47712$$

$$x = \frac{0{,}47712}{0{,}04961} = 9{,}61 \text{ ani.}$$

Rezultă că dacă, în continuare, se va obține același indice mediu de creștere, producția de energie electrică a anului 1975 se va dubla după 6,1 ani și se va tripla după 9,6 ani.

Dacă se pune problema, să se determine cu cît ar crește producția de energie electrică după unul sau două cincinale, păstrîndu-se aceleași ritmuri ale dezvoltării ca și între anii 1965—1975, atunci se va obține :

— după primul cincinal :

$$1{,}121^5 = K$$

$$5 \log 1{,}121 = \log K$$

$$5 \times 0{,}04961 = \log K$$

$$\log K = 0{,}24805$$

$$K = \text{antilog } 0{,}24805 = 1{,}771 \text{ ori sau } 177{,}1\%.$$

Deci, după primul cincinal producția de energie electrică calculată cu un ritm mediu al sporului de 12,1% va fi de 1,771 ori.

— după al doilea cincinal :

$$1{,}121^{10} = K$$

$$10 \log 1{,}121 = \log K$$

$$10 \times 0{,}04961 = \log K$$

$$K = \text{antilog } 0{,}49610 = 3{,}134 \text{ ori.}$$

În sfîrșit, o altă problemă ce trebuie rezolvată este aceea a stabilirii perioadei după care un județ va ajunge din urmă și va depăși un alt județ, ca urmare a faptului că cele două județe au sporuri absolute diferite. Această problemă prezintă un interes practic deosebit, ținînd cont de faptul că în ierarhia județelor pot să apară modificări substanțiale de la o perioadă la alta, că în permanență județele se întrec între ele, în vederea lichidării rămînerii în urmă a unor județe în raport cu altele mai dezvoltate din punct de vedere economic.

Scanned with OKEN Scanner

Pentru rezolvarea acestei probleme este necesar să se cunoască nivelurile absolute pentru momentul de calcul și indicii medii de creștere. Între aceste valori trebuie să existe următoarea relație:

$$Y_A < Y_B \text{ și } \bar{I}_A > \bar{I}_B, \tag{10.72}$$

în care:

Y reprezintă nivelul absolut în momentul de calcul pentru județele A și B;

$\bar{I}$ — indicele mediu anual de creștere.

Deci, se consideră că nivelul absolut al județului A este mai mic decît cel al județului B, dar ritmul mediu de creștere a județului A este mai mare decît ritmul mediu de creștere a județului B.

Problema care se pune este să se găsească numărul de ani după care județul A va ajunge din urmă județul B, respectiv vor avea același nivel absolut al fenomenului studiat, adică:

$$Y'_A = Y'_B, \tag{10.73}$$

în care:

Y' reprezintă nivelul fenomenului în momentul final x.

Aceasta înseamnă că între Y și Y' există relația:

$$Y' = Y_A \cdot \bar{I}_A^x;\ I'_B = Y_B \cdot \bar{I}_B^x. \tag{10.74}$$

Deci, pentru aflarea numărului de ani, după care cele două niveluri absolute diferite și cu ritmuri medii de creștere diferite se vor egala, se cere rezolvarea ecuației:

$$Y_A \cdot \bar{I}_A^x = Y_B \cdot \bar{I}_B^x. \tag{10.75}$$

Pentru exemplificare se vor folosi datele publicate cu privire la producția de oțel a României și Franței pe perioada 1970—1974.

Tabelul 10.20

Volumul și indicii de creștere ai producției de oțel a R.S. România și Franței între anii 1970—1974

Țările	Volumul absolut al indicatorului (mii tone)		Indicele de creștere 1974/1970	Indicele mediu anual 1970—1974
	1970	1974		
0	1	2	3	4
R.S. România	6 517	8 848	1,357	1,079
Franța	773	27 024	1,137	1,033

Sursa: Anuarul Statistic al R.S. România, D.C.S., 1977, p. 561.

Deci, nivelurile inițiale ale indicatorului economic respectiv, pentru care trebuie determinată decalajul, sînt în perioada de calcul (1974):

$$Y_A = 8\,848;\ Y_B = 27\,024,$$

Scanned with OKEN Scanner

iar indicii medii anuali de creştere :

$$\bar{I}_A = 1{,}079\,;\ \bar{I}_B = 1{,}033.$$

În continuare trebuie găsită perioada de timp (x) cînd cele două nivele absolute vor fi egale, adică :

$$Y'_A = Y'_B.$$

În acel moment, cele două nivele absolute vor fi faţă de momentul iniţial ($x = 0$) pentru exemplul considerat :

$$Y'_A = 8\,848 \cdot 1{,}079^x \quad Y'_B = 27\,024 \cdot 1{,}033^x.$$

Ca Y'_A să fie egal cu Y'_B trebuie să se rezolve ecuaţia :

$$8\,848 \cdot 1{,}079^x = 27\,024 \cdot 1{,}033^x,$$

la care se aplică logaritmii :

$$\log 8\,848 + x \log 1{,}079 = \log 27\,024 + x \log 1{,}033\,;$$

$$3{,}94685 + 0{,}03302x = 4{,}43169 + 0{,}01410x\,;$$

$$0{,}01892x = 0{,}48484\,;$$

$$x = \frac{0{,}48484}{0{,}01892} = 25{,}7 \text{ ani.}$$

Deci, după o perioadă de 25,7 ani, ţara noastră va egala Franţa, în ceea ce priveşte producţia totală de oţel, dacă se vor păstra aceleaşi ritmuri medii anuale de creştere. În momentul respectiv producţia celor două ţări va fi de 62 210 mii tone (antilog $[3{,}94\,685 + (25{,}7 \cdot 0{,}03302x)]$, ceea ce va însemna o creştere de 7,03 ori a producţiei ţării noastre $\left(\frac{62\,210}{8\,848}\right)$ şi de 2,302 a producţiei Franţei $\left(\frac{62\,210}{27\,024}\right)$.

Dacă se consideră că timpul de egalare a producţiei este prea mare, se poate modifica ritmul de creştere. De exemplu, dacă ritmul de creştere al României va fi cu un procent mai mare, atunci ecuaţia de rezolvat este :

$$8\,848 \cdot 1{,}089^x = 27\,024 \cdot 1{,}033^x\,;$$

$$\log 8\,848 + x \log 1{,}089 = \log 27\,024 + x \log 1{,}033\,;$$

De unde : $x = 21{,}14$ ani

Scanned with OKEN Scanner

Pentru exemplul luat rezultă că este suficient să se accelereze ritmul producției de oțel a țării noastre cu 1%, iar Franța să-și păstreze ritmul din perioada 1970—1974, pentru ca timpul de egalizare a celor două producții să se reducă cu aproape un cincinal.

În concluzie, analiza seriilor dinamice trebuie să se bazeze pe întregul sistem de indicatori calculați, astfel încît să permită sesizarea tendinței obiective de dezvoltare a fenomenelor cercetate, prin înlăturarea a ceea ce este întîmplător în evoluția fenomenelor studiate.

Pentru aceasta este necesar ca în studiul dinamicii fenomenelor să se folosească serii de date complete, cu indicatorii comparabili între ei și care să se refere la o întreagă etapă de dezvoltare. Numai în măsura în care indicatorii obținuți din prelucrare se bazează pe un număr suficient de date, rezultatele sînt concludente și reflectă veridic evoluția fenomenelor studiate. Această cerință poate fi considerată ca o condiție necesară și suficientă pentru analiza în dinamică a fenomenelor. Numai prin respectarea ei se poate alege corect metoda de ajustare în scopul cunoașterii reale a tendinței de dezvoltare a fenomenelor.

O analiză științifică a evoluției fenomenelor pe o etapă trecută permite calcularea mai departe, a liniei (curbei) de evoluție a fenomenelor și în etapa care urmează, cu alte cuvinte, permite calcularea datelor de perspectivă pentru elaborarea variantelor optime ale planurilor de dezvoltare a economiei naționale.

Scanned with OKEN Scanner

INDICII

Metoda indicilor are o largă aplicabilitate în analiza statistică a variației fenomenelor sociale și economice complexe, pe plan teritorial și în dinamică.

În statistica din țările socialiste, metoda indicilor — datorită avantajelor pe care le prezintă — a fost extinsă și la analiza modului de îndeplinire a planului la nivel micro- și macroeconomic, pe ramuri ale economiei naționale și sub raport teritorial, contribuind la cunoașterea modului de influență a diferiților factori care determină mărimea și dinamica unor fenomene complexe, ce formează obiectul sarcinilor de plan.

Metoda indicilor face parte din cadrul metodelor de analiză factorială, în care variația în timp și spațiu a unui fenomen complex se analizează în funcție de modificarea factorilor săi determinanți. Deci și în acest caz se pune problema să se identifice și să se înregistreze, pe o anumită structură, factorii de care depinde variația unui fenomen complex pentru ca, aplicînd metoda indicilor, să se formeze un sistem de relații care să permită caracterizarea statistică a colectivității în ansamblul ei și a părților sale componente.

Specific metodei indicilor este faptul că fenomenul complex, a cărui variație se studiază, trebuie să se descompună în produsul factorilor înregistrați, printre care există întotdeauna și un factor cantitativ, care, de regulă, are conținut de frecvență de apariție pentru celelalte variabile-factori. Astfel, dacă se notează cu y un fenomen complex și se studiază variația acestuia în timp și spațiu, în funcție de un singur factor x, atunci cel de-al doilea factor de influență îl constituie frecvențele de apariție a diferitelor variante ale variabilei x. Deci, în acest caz, între fenomenul complex și cei doi factori determinanți există relația:

$$y = x \cdot f.$$

Dacă se înregistrează două caracteristici-factori (x, z), atunci fenomenul complex de indexat se descompune în produsul dintre frecvența și valorile perechi ale factorilor înregistrați, adică:

$$y = x \cdot z \cdot f \text{ ș.a.m.d.}$$

De precizat că la aplicarea metodei indicilor se folosesc în culegerea datelor unități complexe de observare, deci caracteristicile-factoriale sînt întotdeauna niveluri parțiale însumate pentru toate unitățile simple ce compun unitatea complexă, sau sînt medii parțiale. Rezultă de aici că atît nivelurile fenomenului complex, cît și nivelurile înregistrate pentru factorii luați în considerare sînt variabile în timp și spațiu în raport cu unitatea complexă care a fost supusă observării statistice.

Pentru a măsura această variație în timp și spațiu a caracteristicilor înregistrate se folosește metoda indicilor.

Indicii se calculează ca raport între nivelurile atinse de un fenomen în două unități diferite de timp sau în două unități diferite de spațiu, sau

Scanned with OKEN Scanner

față de nivelul planificat. Indicii pot fi astfel consideraţi ca mărimi relative de dinamică, de coordonare sau ale planului. Cînd indicii se calculează la nivelul unei grupe sau al întregului ansamblu de unităţi statistice ei au caracter şi de mărime medie, sintetizînd în valoarea lor toate modificările cantitative intervenite în interiorul acestora. În consecinţă, calcularea şi interpretarea lor trebuie să se facă ca pentru orice mărime medie, calculată din mai multe valori individuale diferite.

În aceste condiţii, pentru a asigura un conţinut veridic indicilor statistici este necesar să se ţină seama de următoarele aspecte :

1° datele folosite la calculul indicilor să provină dintr-o observare totală (dări de seamă) sau dacă s-a organizat o cercetare parţială, partea supusă observării să fie reprezentativă în raport cu întreaga colectivitate ;

2° valorile indicilor din numărătorul şi numitorul raportului să fie mărimi reale, sau, dacă sînt calculate folosind o ipoteză statistică, ele să fie practic posibile ;

3° diferenţa dintre numărătorul şi numitorul indicelui să corespundă modificării absolute a fenomenului respectiv.

Reiese că întotdeauna indicii statistici prezintă şi la numărător şi la numitor acelaşi fenomen, dar cu valori diferite, ceea ce face ca indicele folosit să capete denumirea corespunzătoare. De exemplu, dacă se compară cantitatea de produse : „A“ din luna ianuarie a anului curent cu cantitatea de produse „A“ din luna ianuarie a anului precedent, înseamnă că se măsoară variaţia în timp a volumului fizic, deci indicele se va numi „indicele volumului fizic al producţiei“.

În legătură cu folosirea metodei indicilor, mai trebuie adăugat şi faptul că ei se folosesc, de regulă, sub formă de sistem. De exemplu, sistemul indicilor valorii, ai volumului fizic şi ai preţurilor producţiei, ai circulaţiei de mărfuri, sau sistemul indicilor producţiei, productivităţii muncii şi ai cheltuielilor totale de timp de muncă, sau sistemul indicilor recoltei totale, ai recoltei medii şi ai suprafeţelor însămînţate etc. De remarcat că fiecare sistem de indici poate fi considerat ca un sistem independent sau ca un subsistem în cadrul unui sistem mai cuprinzător, fie din punct de vedere organizatoric, fie din punctul de vedere al gradului de cuprindere a caracteristicilor ce se găsesc în relaţii de interdependenţă.

11.1. CLASIFICAREA INDICILOR

Problemele care se pun cu ocazia calculării unor „numere indici“ sînt variate şi soluţionarea lor depinde de conţinutul indicatorilor comparaţi, de natura datelor de care se dispune. Varietatea mare a indicilor întîlniţi în teoria şi practica statistică şi problemele diferite care se pun pentru fiecare formă de indice în parte, necesită o sistematizare a acestora, o clasificare a lor după anumite criterii.

11.1.1. INDICII DINAMICII, TERITORIALI ŞI AI PLANULUI

O primă clasificare a indicilor se face după funcţia pe care o au în studiul variaţiei fenomenelor social-economice. După acest criteriu, indicii

Scanned with OKEN Scanner

se împart în trei grupe mari :

— indici ai dinamicii ;
— indici teritoriali ;
— indici ai planului.

Indicii dinamicii sînt rezultatul comparării în timp a aceluiaşi fenomen y, deci, în acest caz, trebuie să existe o perioadă de bază notată cu y_0 şi mai multe perioade curente care se notează cu $y_1, y_2, \ldots, y_n$.

Indicii teritoriali sînt indicii în care se măsoară variaţia în spaţiu a unui fenomen y în cadrul aceleiaşi perioade de timp. Ei se întîlnesc fie ca mărimi relative de coordonare — cînd se compară două unităţi teritoriale între ele —, fie ca mărimi relative de structură — cînd se compară o unitate teritorială faţă de întreaga ţară. În acest caz trebuie să existe o valoare statistică înregistrată în unitatea teritorială A, notată cu y_A, şi o valoare statistică înregistrată în unitatea teritorială B, notaţă cu y_B. Dacă se compară cu întreaga ţară, atunci intervine în plus valoarea înregistrată pe întregul teritoriu, notată, de obicei, cu y_T.

Indicii planului sînt rezultatul comparării unei valori realizate faţă de valoarea aceluiaşi fenomen înscrisă în plan, sau invers. Dacă se compară nivelul realizat cu nivelul planificat al unui fenomen în cadrul aceleiaşi perioade de timp, atunci se obţine un indice al îndeplinirii planului. Dacă se compară însă nivelul planificat în perioada de plan cu nivelul realizat în perioada de bază, atunci se obţine un indice al sarcinii de plan. Deci, în acest caz, faţă de indicii de dinamică mai intervine în plus nivelul planificat, notat cu y_{pl}.

Problemele metodologice comune tuturor indicilor se vor prezenta pe exemplul indicilor de dinamică şi numai cele specifice vor fi tratate separat şi pentru celelalte două grupe mari : indicii teritoriali şi indicii planului.

11.1.2. Indicii individuali şi de grup

O altă clasificare a indicilor statistici se face după nivelul la care s-au înregistrat datele :

— indici individuali ;
— indici de grup.

Indicii individuali ai unei variabile statistice y se calculează la nivelul unei unităţi de observare şi se notează cu $(i^y_{1/0})$. Indicele individual arată de cîte ori s-a modificat fenomenul y în perioada curentă faţă de perioada de bază. De obicei, indicii (individuali şi de grup) se exprimă sub formă de procente, adică :

$$i^y_{1/0} = \frac{y_1}{y_0} \cdot 100. \qquad (11.1)$$

Indicii de grup ai unei variabile statistice y se notează cu $(I^y_{1/0})$ şi se calculează la nivelul unei grupe sau pe întreaga colectivitate, sintetizînd care este variaţia medie a fenomenului studiat. Deci, *indicele de grup nu*

Scanned with OKEN Scanner

este o sumă a indicilor individuali, ci o medie a acestora, exprimînd tendinţa de modificare în timp şi spaţiu a caracteristicii la care se referă.

De subliniat că elaborarea indicilor de grup ridică cele mai multe probleme care se referă la :

— posibilitatea de însumare a elementelor componente ;

— alegerea bazei de raportare ;

— alegerea formulei de calcul în funcţie de natura datelor de care se dispune ;

— alegerea sistemului de ponderare în funcţie de conţinutul indicatorului de comparat şi relaţiile de sistem utilizate, astfel încît variaţia fenomenului complex să poată fi descompusă în produsul influenţei factorilor care o determină.

Enumerarea acestor probleme arată că ele trebuie să se rezolve, de la caz la caz, pe baza analizei calitative care este obligatorie de fiecare dată cînd se aplică metoda statistică de studiere a fenomenelor de tip colectiv.

Rezultă de aici că, în teoria şi practica statistică, se folosesc mai multe tipuri de indici de grup, care se pot şi ei clasifica după diferite criterii.

11.1.3. Indicii agregaţi, sub formă de medii şi sub formă de raport de medii

După modul de calcul, indicii de grup se pot calcula ca indici agregaţi, ca indici sub formă de medie (aritmetică sau armonică) a indicilor individuali respectivi sau sub formă de raport a două medii.

Indicii sub formă de agregat se calculează raportînd suma nivelurilor parţiale ale unui fenomen din perioada curentă la suma nivelurilor parţiale corespunzătoare din perioada de bază.

De exemplu, dacă se calculează un indice de grup sub formă de agregat pentru fenomenul y, atunci formula de calcul este :

$$I^{y}_{1/0} = \frac{\Sigma y_1}{\Sigma y_0} \cdot 100. \qquad (11.2)$$

În practica statistică se pot întîlni indici agregaţi ale căror elemente sînt însumabile direct, de exemplu, indicele valorii producţiei sau indici agregaţi ai unor fenomene neînsumabile direct, pentru care trebuie găsit un element de echivalenţă, care, în teoria statistică, se mai numeşte şi pondere.

Un indice agregat cu elemente neînsumabile direct este indicele volumului fizic al producţiei. În acest caz se transformă cantităţile vîndute în valori şi apoi se însumează pentru toate produsele. Pentru a nu influenţa valoarea indicelui, se folosesc aceleaşi preţuri şi la numărătorul şi la numitorul indicelui. Aceasta înseamnă că pentru indicele volumului fizic preţurile pot fi considerate ca ponderi.

Indicii sub formă de medie se aplică atunci cînd nu se cunosc toate elementele necesare calculării unui indice agregat. Pentru acesta ei trebuie să îndeplinească următoarele două condiţii :

1° să fie o medie a indicilor individuali ai fenomenului de indexat ;

2° să fie egali în valoare cu indicele agregat pe care îl substituie.

Scanned with OKEN Scanner

Pentru a respecta aceste condiţii trebuie să se folosească fie o medie aritmetică, fie o medie armonică a indicilor individuali, la care ponderea să se facă cu unul din elementele ce formează indicele agregat.

Indicele agregat, fiind o mărime relativă generală cu conţinut de mărime medie a unor mărimi relative parţiale, înseamnă că se poate transforma fie într-o medie aritmetică a indicilor individuali ponderaţi cu valorile din numitorii din care s-au calculat, fie ca o medie armonică la care ponderea să se facă cu numărătorii parţiali, adică :

— indicele de forma unei medii aritmetice :

$$I^{v}_{1/0} = \frac{\Sigma i^{v}_{1/0} y_0}{\Sigma y_0}; \qquad (11.3)$$

— indicele de forma unei medii armonice :

$$I^{v}_{1/0} = \frac{\Sigma y_1}{\sum \frac{1}{i^{v}_{1/0}} y_1}. \qquad (11.4)$$

Este uşor de demonstrat că cei doi indici sub formă de medie sînt egali cu indicele agregat al variabilei y şi, deci, egali între ei.

În cazul mediei aritmetice, dacă se înlocuiesc indicii individuali cu formula lor de calcul se obţine :

$$I^{v}_{1/0} = \frac{\Sigma i^{v}_{1/0}\, y_0}{\Sigma y_0} = \frac{\sum \frac{y_1}{y_0} y_0}{\Sigma y_0} = \frac{\Sigma y_1}{\Sigma y_0}.$$

În cazul indicelui de forma mediei armonice se procedează la fel, adică :

$$I^{v}_{1/0} = \frac{\Sigma y_1}{\sum \frac{1}{i^{v}_{1/0}} y_1} = \frac{\Sigma y_1}{\sum \frac{1}{\frac{y_1}{y_0}} y_0} = \frac{\Sigma y_1}{\sum \frac{y_0}{y_1} y_1} = \frac{\Sigma y_1}{\Sigma y_0}.$$

Rezultă de aici că, dacă se respectă condiţiile de folosire a mediei aritmetice şi armonice pentru calculul mediei mărimilor relative parţiale, indicele aritmetic se foloseşte cînd se cunosc indicii individuali şi valorile din numitorii parţiali, iar indicele armonic cînd se cunosc indicii individuali şi valorile din numărătorii parţiali.

De subliniat că indicii sub formă de medie aritmetică se folosesc şi în cazul cînd datele provin dintr-o observare selectivă şi se pune problema de a măsura eroarea medie şi eroarea limită de reprezentativitate. În acest sens, indicele mediu aritmetic se foloseşte ca orice valoare medie pentru

Scanned with OKEN Scanner

care i se poate calcula dispersia, interpretîndu-se, în consecinţă, la estimarea indicilor din colectivitatea generală.

Indicii sub formă de raport a două medii se întîlnesc frecvent în teoria şi practica statistică pentru măsurarea variaţiei unor caracteristici derivate ce se formează ca mărime medie la nivelul unei grupe sau pe întreaga colectivitate.

În formă generală, indicele sub formă de raport a două medii se prezintă astfel :

$$I_{1/0}^{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0}. \tag{11.5}$$

Într-o formă explicită, acest indice se poate scrie astfel încît să apară distinct că valoarea unei medii depinde de valorile individuale ale caracteristicii x şi de frecvenţele de apariţie a acestora (f), adică :

$$I_{1/0}^{\bar{x}} = \frac{\Sigma x_1 f_1}{\Sigma f_1} : \frac{\Sigma x_0 f_0}{\Sigma f_0}. \tag{11.6}$$

După cum se ştie, la calculul celor două medii se pot folosi frecvenţele absolute (f_i) sau frecvenţele relative (f_i^*). În cazul folosirii frecvenţelor relative luate în procente, indicele sub forma de raport a două medii se poate scrie :

$$I_{1/0}^{\bar{x}} = \frac{\Sigma x_1 f_1^* \%}{100} : \frac{\Sigma x_0 f_0^* \%}{100} = \frac{\Sigma x_1 f_1^* \%}{\Sigma x_0 f_0^* \%}. \tag{11.7}$$

Deci, în acest caz, indicele raportului a două medii se transformă într-un indice agregat al fenomenului complex ce rezultă din produsul respectiv de frecvenţă. De exemplu, dacă variabila x reprezintă productivitatea muncii, numărul de muncitori ce revine pentru fiecare nivel de productivitate a muncii formează frecvenţele f, atunci, folosind relaţia (11.7), se obţine aceeaşi valoare ca şi cînd s-ar calcula indicele dinamicii volumului fizic al producţiei doar pe seama modificării productivităţii muncii.

Analizînd relaţia (11.7) reiese că factorul cantitativ influenţează modificarea în timp a valorii medii a unei caracteristici statistice prin intermediul modificării de structură. Dacă între cele două perioade comparate există diferenţe de structură, ele vor influenţa în sensul măririi mediei, atunci cînd predomină tendinţa de creştere a ponderii valorilor mai mari — sau în sensul scăderii mediei — dacă predomină tendinţa de creştere a ponderii valorilor mai mici ale caracteristicii.

11.1.4. Indicii cu structură variabilă, cu structură fixă şi de variaţie a structurii

În funcţie de structura la care se calculează variaţia unei caracteristici, indicii sub formă de raport a două medii pot fi :

— indici cu structură variabilă ;

Scanned with OKEN Scanner

— indici cu structură fixă;
— indici ai variației structurii, denumiți și indici ai modificărilor structurale.

Indicele cu structură variabilă se folosește cînd se compară două valori medii ale aceleiași caracteristici calculate pe baza valorilor individuale (factor calitativ) în funcție de frecvențele lor de apariție (factor cantitativ). Acest indice se calculează folosind relațiile (11.6) și (11.7), după cum se cunosc frecvențele absolute sau frecvențele relative.

Indicele de structură fixă se folosește atunci cînd se pune problema să se măsoare numai influența variației valorilor individuale din care s-au calculat mediile în cele două perioade, analizîndu-se influența modificărilor din structura colectivității.

Din punct de vedere matematic, la calculul indicelui cu structură fixă se pot folosi atît ponderile din perioada de bază, cît și ponderile din perioada curentă, alegerea unei variante sau a celeilalte făcîndu-se în funcție de scopul cercetării.

Indicele cu structură fixă, folosind ponderile perioadei de bază, corespunde ipotezei ca frecvențele perioadei curente să fie egale cu cele din perioada de bază. În acest caz, formula de calcul va fi :

$$I_{1/0}^{\bar{x}(x)} = \frac{\sum x_1 f_0}{\sum f_0} : \frac{\sum x_0 f_0}{\sum f_0} \qquad (11.8)$$

$$f_1 = f_0.$$

Indicele cu structură fixă, folosind ponderile perioadei curente, corespunde ipotezei ca variația caracteristicii x să se studieze în funcție de forma de distribuție a frecvențelor din perioada curentă. Formula de calcul a indicelui cu structură fixă va fi :

$$I_{1/0}^{\bar{x}(x)} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1}; \qquad (11.9)$$

$$f_0 = f_1.$$

Particularizarea unei formule sau a celeilalte se face de la caz la caz, în raport cu sistemul de ponderare adoptat și de perioada la care se referă datele.

În literatura de specialitate, cei mai mulți autori optează pentru ponderarea cu factorul cantitativ din perioada curentă pentru măsurarea gradului de influență al factorului calitativ (11.9).

Pornind de la relațiile (11.8) și (11.9) se poate trece la indicele agregat al fenomenului complex y, pentru care media reprezintă unul din factorii determinanți (11.10) și (11.11).

În cazul ponderării cu factorul cantitativ din perioada de bază se obține :

$$I_{1/0}^{\bar{x}(x)} = \frac{\sum x_1 f_0}{\sum f_0} : \frac{\sum x_0 f_0}{\sum f_0} = \frac{\sum x_1 f_0}{\sum x_0 f_0}, \qquad (11.10)$$

$$f_1 = f_0$$

Scanned with OKEN Scanner

deci :

$$I_{1/0}^{y(x)} = \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0};$$

$$f_1 = f_0$$

în care :

x şi f sînt factorii de care depinde variaţia fenomenului complex y.

În cazul ponderării cu factorul cantitativ din perioada curentă se obţine :

$$I_{1/0}^{\bar{x}(x)} = \frac{\Sigma x_1 f_1}{\Sigma f_1} : \frac{\Sigma x_0 f_1}{\Sigma f_1} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1}; \qquad (11.11)$$

$$f_0 = f_1.$$

Oricare din formulele folosite la calculul indicilor cu structură fixă presupune cunoaşterea separată a valorilor caracteristicii x şi a frecvenţelor din ambele perioade. În practica statistică se întîlnesc însă cazuri cînd nu se dispune de toate elementele şi atunci indicele agregat se transformă într-un indice mediu al indicilor individuali ai variabilei x.

Analizînd formulele de calcul ale indicilor agregaţi, se alege tipul de medie ce se va utiliza ţinînd seama ca ponderile folosite să fie acelea care nu necesită calcule suplimentare, adică ambele elemente să se refere la aceeaşi perioadă. Astfel, dacă se porneşte de la formula (11.10) atunci ponderea care se alege este aceea din numitor, deci se va folosi media aritmetică a indicilor individuali, adică :

$$I_{1/0}^{\bar{x}(x)} = \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0} = \frac{\Sigma i_{1/0}^{x} x_0 f_0}{\Sigma x_0 f_0}; \qquad (11.12)$$

$$f_0 = f_1.$$

În cazul formulei (11.11), în care ponderea se face cu factorul cantitativ din perioada curentă, convenabil este să se folosească media armonică, deoarece elementele din aceeaşi perioadă sînt la numărător :

$$I_{1/0}^{\bar{x}(x)} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1} = \frac{\Sigma x_1 f_1}{\Sigma \frac{1}{i_{1/0}^{x}} x_1 f_1}; \qquad (11.13)$$

$$f_0 = f_1.$$

În concluzie, se poate spune că, în practica statistică, pentru indicele cu structură fixă se pot întîlni trei variante de calcul :

— prima variantă, în care indicele se calculează ca raport a două medii condiţionate de aceeaşi distribuţie de frecvenţe (11.8) şi (11.9);

— a doua variantă, în care se foloseşte un indice de formă agregată cu factorul calitativ variabil şi factorul cantitativ invariabil de la o perioadă la alta (11.10) şi (11.11);

Scanned with OKEN Scanner

— a treia variantă, în care se foloseşte o medie aritmetică sau armonică a indicilor individuali (11.12) şi (11.13).

Alegerea variantei de calcul se face în funcţie de datele de care se dispune şi de scopul analizei.

Indicele de variaţie a structurii colectivităţii, denumit şi indicele modificărilor structurale, se foloseşte pentru a stabili numai influenţa distribuţiei factorului cantitativ în condiţiile în care valorile individuale ar fi rămas neschimbate. Şi aici se pot construi două variante de calcul după ponderile utilizate. În mod unanim s-a adoptat formula de calcul în care ponderea se face cu perioada de bază. Indicele modificărilor structurale se calculează după formula:

$$I_{1/0}^{\bar{x}(f)} = \frac{\Sigma x_0 f_1}{\Sigma f_1} : \frac{\Sigma x_0 f_0}{\Sigma f_0}; \qquad (11.14)$$

$$x_1 = x_0.$$

Indicele modificărilor structurale şi indicele cu structură fixă apar ca indici factoriali ai indicelui cu structură variabilă. Aceşti trei indici formează împreună un sistem în cadrul căruia se pot stabili o serie de relaţii bine determinate.

Dacă aceşti indici se încadrează într-un sistem mai larg, în care variaţia unui indicator absolut este luată în funcţie de variaţia unei valori medii, atunci şi indicii agregaţi respectivi pot fi consideraţi în raport cu structura lor ca fiind cu structură variabilă, cu structură fixă şi de variaţie a structurii colectivităţii.

De exemplu, dacă caracteristica este un indicator absolut — volumul fizic al producţiei — care depinde de valoarea medie a unei caracteristici x — productivitatea muncii — şi de frecvenţele de apariţie a acestora — numărul de muncitori —, atunci sistemul de indici va fi:

— indicele variabilei y care, în raport cu x şi f, este un indice cu structură variabilă:

$$I_{1/0}^{y} = \frac{\Sigma y_1}{\Sigma y_0} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_0}; \qquad (11.15)$$

— indicele variabilei y, calculat numai sub influenţa factorului x, care poate fi considerat ca un indice cu structură fixă şi care, de regulă, apare ponderat cu frecvenţele din perioada curentă:

$$I_{1/0}^{y(x)} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1}; \qquad (11.16)$$

— indicele variabilei y, calculat numai în funcţie de modificarea factorului cantitativ, care, după cum s-a mai arătat, exprimă modificarea din structura colectivităţii pentru care se înregistrează datele, adică:

$$I_{1/0}^{y(f)} = \frac{\Sigma x_0 f_1}{\Sigma x_0 f_0}. \qquad (11.17)$$

Scanned with OKEN Scanner

11.1.5. Indicii cu bază fixă şi cu bază variabilă

După baza de raportare, indicii dinamicii fenomenelor pot fi indici cu bază fixă şi indici cu bază variabilă.

Indicii cu bază fixă sînt aceia în care variaţia unei caracteristici se măsoară faţă de o singură bază de referinţă.

Indicii cu bază variabilă sînt aceia în care variaţia caracteristicii se măsoară succesiv faţă de perioada precedentă sau în cadrul aceleiaşi perioade faţă de o altă unitate teritorială.

Indicii cu bază fixă şi cu bază variabilă se calculează atît ca indici individuali[1], cît şi ca indici de grup. Probleme deosebite apar pentru indicii de grup cu bază fixă sau variabilă a unei caracteristici neînsumabilă direct, în legătură cu alegerea ponderilor.

11.1.6. Indicii cu ponderi constante şi ponderi variabile

După ponderile folosite, într-o serie de indici se pot întîlni:
— indici cu ponderi constante;
— indici cu ponderi variabile.

Indicii cu ponderi constante sînt indicii în care, pentru toată seria, se folosesc aceleaşi ponderi.

De exemplu, dacă se analizează variaţia caracteristicii x pentru o perioadă de cinci ani, indicii de grup de formă agregată se pot calcula cu bază fixă şi cu bază variabilă în care ponderile folosite, notate cu f, să fie aceleaşi pentru toată seria de indici. Pentru aceasta este necesar să se dispună de mărimile absolute ale caracteristicii x şi de ponderile f ale acesteia, pe baza cărora se vor elabora cele două serii cronologice iniţiale:

$$x_0;\ x_1;\ x_2;\ x_3;\ x_4;$$

$$f_0;\ f_1;\ f_2;\ f_3;\ f_4.$$

Combinînd aceste elemente, se vor putea calcula mai multe serii de indici, care vor diferi prin baza de raportare şi ponderile utilizate. Alegerea ponderilor pentru o serie de indici se face în funcţie de conţinutul indicatorului a cărui variaţie se studiază. În unele cazuri, ponderile sînt alese din perioada de bază a indicelui (de exemplu, la indicele volumului fizic); în alte cazuri ponderile sînt alese din perioada curentă (de exemplu, la indicele preţurilor).

În cazul folosirii *ponderilor constante* luate din perioada de bază a întregii serii (f_0) se obţin următorii *indici cu bază fixă*:

$$I^x_{1/0} = \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0};\qquad I^x_{2/0} = \frac{\Sigma x_2 f_0}{\Sigma x_0 f_0};\qquad (11.18)$$

$$I^x_{3/0} = \frac{\Sigma x_3 f_0}{\Sigma x_0 f_0};\qquad I^x_{4/0} = \frac{\Sigma x_4 f_0}{\Sigma x_0 f_0}.$$

[1] Aceste probleme s-au tratat în cadrul capitolului „Analiza seriilor cronologice" la indicatorii relativi, unde s-au arătat şi relaţiile dintre ei.

Scanned with OKEN Scanner

Formula generalizată a acestei serii de indici va fi:

$$I^x_{i/0} = \frac{\sum x_i f_0}{\sum x_0 f_0} \; ; \; i = 1, 2, 3, \ldots, n. \tag{11.19}$$

În mod corespunzător se obţin şi *indici cu bază în lanţ* şi *cu ponderi constante* luate din perioada de bază a seriei:

$$I^x_{1/0} = \frac{\sum x_1 f_0}{\sum x_0 f_0}; \; I^x_{2/1} = \frac{\sum x_2 f_0}{\sum x_1 f_0}; \; I^x_{3/2} = \frac{\sum x_3 f_0}{\sum x_2 f_0}. \tag{11.20}$$

Formula generalizată pentru această serie de indici este:

$$I^x_{i/i-1} = \frac{\sum x_i f_0}{\sum x_{i-1} f_0} \quad i = 1, 2, 3, \ldots, n. \tag{11.21}$$

La calculul aceloraşi indici se pot folosi ponderile constante din perioada curentă a seriei (pentru exemplul luat ponderea constantă din perioada curentă a seriei este f_4).

Indicii cu bază fixă şi cu ponderea constantă din perioada curentă a indicelui vor fi:

$$I^x_{1/0} = \frac{\sum x_1 f_4}{\sum x_0 f_4}; \; I^x_{2/0} = \frac{\sum x_2 f_4}{\sum x_0 f_4}; \tag{11.22}$$

$$I^x_{3/0} = \frac{\sum x_3 f_4}{\sum x_0 f_4}; \; I^x_{n/0} = \frac{\sum x_n f_4}{\sum x_0 f_4}.$$

Formula generalizată a acestor indici va fi:

$$I^x_{i/0} = \frac{\sum x_i f_n}{\sum x_0 f_n} \quad i = 1, 2, 3, \ldots, n. \tag{11.23}$$

Indicii cu baza în lanţ şi cu ponderile perioadei curente a seriei:

$$I^x_{1/0} = \frac{\sum x_1 f_4}{\sum x_0 f_4}; \; I^x_{2/1} = \frac{\sum x_2 f_4}{\sum x_1 f_4}; \tag{11.24}$$

$$I^x_{3/2} = \frac{\sum x_3 f_4}{\sum x_2 f_4}; \; I^x_{4/3} = \frac{\sum x_4 f_4}{\sum x_3 f_4}.$$

Formula generalizată a acestor indici va fi:

$$I^x_{i/i-1} = \frac{\sum x_i f_n}{\sum x_{i-1} f_n} \quad i = 1, 2, 3, \ldots, n. \tag{11.25}$$

Scanned with OKEN Scanner

De subliniat că relaţiile care există între indicii cu bază fixă şi cei cu baza în lanţ, prezentate la capitolul „Analiza seriilor cronologice", se verifică numai în cazul folosirii ponderilor constante. De exemplu, dacă se înmulţesc succesiv trei indici cu baza în lanţ şi cu ponderi constante se obţine indicele cu bază fixă pe întreaga perioadă.

În cazul ponderării cu perioada de bază a seriei :

$$\frac{\sum x_1 f_0}{\sum x_0 f_0} \cdot \frac{\sum x_2 f_0}{\sum x_1 f_0} \cdot \frac{\sum x_3 f_0}{\sum x_2 f_0} = \frac{\sum x_3 f_0}{\sum x_0 f_0} = I^x_{3/0}. \qquad (11.26)$$

Sau în cazul ponderării cu perioada curentă a seriei :

$$\frac{\sum x_1 f_3}{\sum x_0 f_3} \cdot \frac{\sum x_2 f_3}{\sum x_1 f_3} \cdot \frac{\sum x_3 f_3}{\sum x_2 f_3} = \frac{\sum x_3 f_3}{\sum x_0 f_3} = I^x_{3/0} \qquad (11.27)$$

Toţi aceşti indici, calculaţi cu ponderi constante luate fie din perioada curentă a seriei, fie din perioada de bază a seriei, prezintă avantajul că elimină influenţa modificării ponderilor de la o unitate de timp la alta. Pentru o perioadă foarte mare de timp, aceşti indici nu sînt întotdeauna suficient de concludenţi, deoarece pot să apară modificări esenţiale în structura ponderilor. De exemplu, pentru indicii în care preţul este luat ca pondere se construiesc mai multe serii de indici, calculaţi cu preţurile comparabile care se stabilesc periodic în funcţie de modificările intervenite în condiţiile de formare a lor. Trecerea de la o serie de indici la alta se face numai după ce s-a stabilit un indice de modificare a preţurilor comparabile cu care se caracterizează indicii calculaţi iniţial.

În sfîrşit, în analiza bazată pe aplicarea metodei indicilor se pune problema ca, în paralel, să se calculeze şi indicii cu ponderi variabile.

Indicii cu ponderi variabile sînt aceia în care ponderea se schimbă odată cu schimbarea bazei de raportare. De regulă, aceşti indici se calculează numai cu baza în lanţ, folosind fie ponderea din perioada curentă, fie pe cea din perioada de bază a indicelui respectiv.

În cazul folosirii *ponderilor variabile* luate din perioada de bază a *indicelui cu baza în lanţ* se obţine următoarea serie de indici :

$$I^x_{1/0} = \frac{\sum x_1 f_0}{\sum x_0 f_0}\ ;\ I^x_{2/1} = \frac{\sum x_2 f_1}{\sum x_1 f_1}\ ;$$

$$I^x_{3/2} = \frac{\sum x_3 f_2}{\sum x_2 f_2}\ ;\ I^x_{4/3} = \frac{\sum x_4 f_3}{\sum x_3 f_3}. \qquad (11.28)$$

Formula generală de calcul va fi :

$$I^x_{i/i-1} = \frac{\sum x_i f_{i-1}}{\sum x_{i-1} f_{i-1}}. \qquad (11.29)$$

Scanned with OKEN Scanner

În cazul folosirii *ponderilor variabile* luate din *perioada curentă a indicelui cu baza în lanț* se obține seria de indici:

$$I^x_{1/0} = \frac{\sum x_1 f_1}{\sum x_0 f_1}\ ;\ I^x_{2/1} = \frac{\sum x_2 f_2}{\sum x_1 f_2}\ ;$$
$$I^x_{3/2} = \frac{\sum x_3 f_3}{\sum x_2 f_3}\ ;\ I^x_{4/3} = \frac{\sum x_4 f_4}{\sum x_3 f_4}\ . \qquad (11.30)$$

Formula generală de calcul va fi:

$$I^x_{i/i-1} = \frac{\sum x_i f_i}{\sum x_{i-1} f_i}\ . \qquad (11.31)$$

În cazul indicilor cu ponderi variabile, trecerea de la indicii cu baza în lanț la cei cu bază fixă nu se mai poate face, deoarece produsul indicilor cu baza în lanț și cu ponderi variabile nu este egal cu indicele cu baza fixă al întregii perioade:

$$\frac{\sum x_1 f_1}{\sum x_0 f_1} \cdot \frac{\sum x_2 f_2}{\sum x_1 f_2} \cdot \frac{\sum x_3 f_3}{\sum x_2 f_3} \neq \frac{\sum x_3 f_3}{\sum x_0 f_3}\ . \qquad (11.32)$$

După cum reiese, din cele prezentate pînă aici, se pot obține foarte multe variante de indici în funcție de baza de raportare și ponderea utilizată. Alegerea variantelor de calcul pentru o serie de indici de dinamică se face de la caz la caz, în așa fel încît să se asigure un conținut obiectiv valorii statistice calculate.

11.2. SISTEME DE PONDERARE FOLOSITE LA CONSTRUIREA INDICILOR. RELAȚIA LUI BORTKIEWICZ

În practica statistică problemele cele mai dificile apar în legătură cu alegerea și folosirea ponderilor la elaborarea indicilor de grup.

Alegerea și folosirea sistemelor de ponderare trebuie să se facă în mod diferențiat, ținînd seama de conținutul indicatorului de comparat, de natura datelor existente în evidența curentă și de posibilitatea de a stabili o analogie între descompunerea pe factori a modificării absolute și relative.

În cursul dezvoltării statisticii s-au propus mai multe sisteme de ponderare, care au fost particularizate, de regulă, pe exemplul indicelui volumului fizic sau al prețurilor producției și circulației de mărfuri. Pentru generalizare, însă, se vor prezenta sistemele de indici folosind o variabilă complexă y, dependentă de un factor calitativ x și de un factor cantitativ f.

Avînd trei variabile înregistrate la nivelul unităților complexe, care formează în mod permanent colectivitatea supusă observării, înseamnă că se pot calcula trei indici individuali și trei indici de grup.

Scanned with OKEN Scanner

Indicii individuali se calculează ca indici simpli folosind datele înregistrate pentru fiecare variabilă la nivelul unității de observare folosită.

Indicii de grup, calculați la nivelul întregului ansamblu, se calculează ca indici agregați. Pentru prezentarea diferitelor sisteme de ponderare se presupune că variabila y este însumabilă direct și că se descompune integral la nivelul fiecărei unități în produsul dintre variabila x — cu caracter de mărime statistică derivată — și variabila f — cu caracter de variabilă cantitativă neînsumabilă direct. Aceasta înseamnă că pentru indicii de grup ai celor două variabile — factori se va folosi cîte un indice agregat în care, succesiv, fenomenul de indexat este variabil, iar celălalt are caracter de pondere.

Un prim sistem de ponderare este acela propus în 1864 de statisticianul belgian *Laspeyres*, la care ponderile folosite sînt cele din perioada de bază. În acest caz se obține:

— indicele unei variabile calitative x, de tip *Laspeyres*, este:

$$I^{x}_{1/0} = \frac{\sum x_1 f_0}{\sum x_0 f_0}; \tag{11.33}$$

— indicele unei variabile cantitative f, de tip *Laspeyres*, este:

$$I^{f}_{1/0} = \frac{\sum f_1 x_0}{\sum f_0 x_0}. \tag{11.34}$$

Adoptarea sau respingerea celor doi indici nu se poate face decît după ce se analizează conținutul lor și măsura în care ei reflectă niște proporții reale cu privire la dezvoltarea fenomenelor la care se referă. În consecință, fiecare indice trebuie analizat separat, corespunzător cu conținutul indicatorilor absoluți pe care îi conține și cu relațiile de interdependență dintre fenomenul de indexat și ponderile folosite.

După cum s-a mai arătat, în mod frecvent în teoria și practica statistică se întîlnesc indici ai unor caracteristici cantitative care se obțin prin centralizarea unităților simple la nivelul unităților complexe de observare și indici ai unor caracteristici calitative care se exprimă fie printr-un raport, fie printr-o mărime medie și care, de regulă, au caracter de atribut al factorului cantitativ (de exemplu, productivitatea este un atribut al muncitorilor ocupați în sfera producției materiale etc.).

În primul caz, valorile calculate la nivelul unităților de observare folosite sînt independente de valorile corespunzătoare ale elementelor care se folosesc ca ponderi. De aceea, pentru astfel de caracteristici cantitative se folosesc, de obicei, ponderile perioadei de bază corespunzătoare ipotezei în care s-ar fi modificat numai factorul cantitativ și factorul calitativ care joacă rol de pondere ar fi rămas la același nivel de dezvoltare. În schimb, la măsurarea dinamicii fenomenului calitativ, dacă se folosesc ponderile perioadei de bază, nu s-ar ține seama de dependența sa față de modificările factorului cantitativ la care se referă.

În concluzie, în teoria și practica statistică, se apreciază că numai indicele factorului cantitativ se poate calcula după sistemul de ponderare folosit de Laspeyres. Indicele unei caracteristici calitative de tip Laspeyres

Scanned with OKEN Scanner

se consideră că nu este suficient de semnificativ, deoarece nu ține seama tocmai de variația produsă prin dependența sa față de structura colectivității.

Un alt sistem de ponderare este cel propus în 1874 de către statisticianul german *Paasche*, la care ponderile utilizate sînt cele din perioada curentă a indicelui:

— pentru indicele factorului cantitativ, formula va fi:

$$I^{f}_{1/0} = \frac{\sum f_1 x_1}{\sum f_0 x_1}; \tag{11.35}$$

— pentru indicele factorului calitativ, formula va fi:

$$I^{x}_{1/0} = \frac{\sum x_1 f_1}{\sum x_0 f_1}. \tag{11.36}$$

Și aceste formule trebuie să fie analizate în raport cu conținutul și scopul analizei. Pe lîngă argumentele legate de conținutul fenomenelor de indexat, și care au fost prezentate la interpretarea indicilor de tip Laspeyres, se mai adaugă și unele aspecte legate de conținutul și valoarea ponderilor. Acceptarea sau respingerea sistemului de ponderare trebuie să se facă și în funcție de tendința obiectivă de modificare a prețurilor. Dacă tendința obiectivă a fenomenului este de scădere, atunci ponderile perioadei curente utilizate vor influența în sensul diminuării indicelui — ceea ce este nejustificat pentru cazul în care evoluția caracteristicii de indexat este independentă de variația ponderilor.

În teoria indicilor statistici se mai întîlnesc și alte sisteme de ponderare, care țin seama de ponderile din ambele perioade de calcul.

Indicele prețurilor, calculat de *Edgeworth*, se bazează pe cumularea cantităților din perioada de bază cu cele din perioada curentă, pe care le folosește drept ponderi pentru variația prețurilor. Prețul, fiind o variabilă calitativă, formula poate fi generalizată astfel:

$$I^{x}_{1/0} = \frac{\sum x_1(f_1 + f_0)}{\sum x_0(f_1 + f_0)}. \tag{11.37}$$

Acest indice prezintă dezavantajul principal că poate fi particularizat numai pentru variația unui factor calitativ, iar ponderea este un factor cantitativ care poate fi însumat de la o perioadă la alta. Pentru analiza variației unui factor cantitativ — de exemplu pentru indicele volumului fizic al producției — nu se poate aplica, deoarece aici se folosesc ca ponderi valorile unui factor calitativ de forma unor mărimi relative sau medii care nu au sens să fie cumulate de la o perioadă la alta. Neputîndu-se extinde și la analiza variației factorului cantitativ, acest indice nu poate fi inclus într-un sistem de indici în cadrul căruia să se poată stabili gradul de influență a diferiților factori asupra fenomenului complex pe care îl determină.

Scanned with OKEN Scanner

Un alt indice bazat pe folosirea ponderilor din ambele perioade este indicele „ideal" al lui *Fisher*. El a calculat un indice de grup al prețurilor ca o medie geometrică a celor doi indici agregați de tip Laspeyres și de tip Paasche.

Generalizînd formulele lui Fisher, pentru orice caracteristică calitativă se obține :

$$I^{x}_{1/0} = \sqrt{\frac{\sum x_1 f_0}{\sum x_0 f_0} \cdot \frac{\sum x_1 f_1}{\sum x_0 f_1}}. \qquad (11.38)$$

Acest indice prezintă avantajul că se încadrează în intervalul de variație a valorilor indicilor calculați pe baza celor două sisteme de ponderare, deci va compensa o parte din tendința de modificare a ponderilor folosite. Din punct de vedere practic prezintă însă dezavantajul că necesită cunoașterea separată a tuturor elementelor de calcul și combinarea tuturor variantelor posibile. În practică este foarte greu să se dispună de toate aceste elemente, mai ales dacă sfera de cuprindere a colectivității la care se referă sistemul de indici este foarte mare, ca, de exemplu, o ramură de producție sau întreaga economie națională.

Indicele lui Fisher se poate aplica pentru orice caracteristică statistică a cărei variație se măsoară cu ajutorul unui indice ponderat. În consecință, indicele factorului cantitativ de tip Fisher este :

$$I^{f}_{1/0} = \sqrt{\frac{\sum x_0 f_1}{\sum x_0 f_0} \cdot \frac{\sum x_1 f_1}{\sum x_1 f_0}}. \qquad (11.39)$$

Indicele lui Fisher se folosește, în special, la calculul indicilor teritoriali pe plan internațional, de exemplu, la calculul indicilor consumului populației, unde prețurile luate ca ponderi se referă la ambele țări comparate.

Din cele prezentate pînă aici rezultă că indicii se pot construi cu sisteme diferite de ponderare, care se referă însă fie la perioada de bază, fie la perioada curentă.

Pentru a analiza gradul și sensul de influență ale ponderilor se folosește, în teoria și practica indicilor statistici, *relația lui Bortkiewicz*. Aplicată la indicii unui factor calitativ se obține :

$$I^{x(f_1)}_{1/0} : I^{x(f_0)}_{1/0} = 1 + r_{1}^{x}{}_{1}^{f} \cdot V_1^x \cdot V_1^f, \qquad (11.40)$$

$$I^{x(f_1)}_{1/0} = \frac{\sum x_1 f_1}{\sum x_0 f_1} \; ; \; I^{x(f_0)}_{1/0} = \frac{\sum x_1 f_0}{\sum x_0 f_0}$$

în care :

$r_{1}^{x}{}_{1}^{f}$ reprezintă coeficientul de corelație liniară simplă dintre indicii individuali și factorului calitativ $_1^x$ și indicii individuali ai ponderilor, respectiv indicii individuali ai factorului cantitativ $_1^f$;

V_1^x — coeficientul de variație a indicilor individuali ai variabilei x față de indicele lor mediu ;

V_1^f — coeficientul de variație a indicilor individuali ai ponderilor față de indicele lor mediu.

Scanned with OKEN Scanner

Interpretînd această relaţie rezultă că sistemul de ponderare nu influenţează asupra valorii indicelui de grup al unei variabile calitative exprimată însă numeric, dacă produsul celor trei factori este nul, adică :

$$r_i^{x,f} \cdot V_i^x \cdot V_i^f = 0. \qquad (11.41)$$

Pentru ca acest produs de trei factori să fie nul, trebuie ca cel puţin unul din factori să fie egal cu zero :

— primul factor :

$$r_i^{x,f} = 0. \qquad (11.42)$$

Coeficientul de corelaţie liniară simplă este egal cu zero cînd cele două variabile sînt independente între ele. Aceasta este posibil în cazul în care modificările individuale ale factorului calitativ nu se asociază cu modificarea factorului cantitativ deci nu există legătură statistică ;

— al doilea factor :

$$V_i^f = 0 \qquad (11.43)$$

Pentru ca valoarea coeficientului de variaţie să fie nulă, trebuie ca valorile din care s-a calculat media respectivă să fie egale între ele. Deci, coeficientul de variaţie al factorului calitativ este egal cu zero cînd toate valorile sale individuale înregistrate s-au modificat în acelaşi sens şi în aceeaşi proporţie. În acest caz, indicele pe total este egal cu oricare din indicii individuali ai variabilei calitative ;

— al treilea factor :

$$V_i^x = 0. \qquad (11.44)$$

Coeficientul de variaţie al indicilor factorului cantitativ este egal cu zero cînd valorile individuale ale acestuia s-au modificat în mod constant pentru toate unităţile de observare folosite. În această situaţie indicii individuali sînt egali între ei şi deci egali cu indicele de grup al acestora.

Toate aceste trei variante constituie însă nişte cazuri limită, care se întîlnesc extrem de rar în practica statistică. De regulă, cei trei factori sînt diferiţi de zero şi, după modul în care ei se asociază, pot influenţa ca sens şi mărime valoarea indicilor calculaţi cu sisteme de ponderare diferite.

În cazul în care produsul celor trei factori este diferit de zero, deci fiecare are o valoare semnificativă, sînt două variante :

— prima variantă :

$$r_i^{f,f} \cdot V_i^x \cdot V_i^f > 0. \qquad (11.45)$$

Pentru ca acest produs să fie pozitiv, trebuie ca cei trei factori să fie mai mari decît zero. Semnul produsului celor trei factori depinde de sensul coeficientului de corelaţie liniară simplă, deoarece ceilalţi doi, calculîndu-se pe baza raportului dintre abaterea medie pătratică şi media caracteristicii, sînt întotdeauna pozitivi. Rezultă că atunci cînd între cele două serii de indici individuali (ai variabilei calitative şi ai ponderilor) există o corelaţie directă se verifică relaţia (11.45) :

— a doua variantă :

$$r_i^{x,f} \cdot V_i^x \cdot V_i^f < 0. \qquad (11.46)$$

Scanned with OKEN Scanner

Pentru ca produsul celor trei factori să fie negativ — deci să verifice relaţia (11.46) — trebuie ca între cele două serii de indici individuali ($_i^x$, $_i^f$) să fie o corelaţie inversă.

În concluzie, interpretarea influenţei sistemelor de ponderare folosite asupra valorii unui indice de grup se bazează pe următoarele trei cazuri:

1° — indicele de grup, calculat cu ponderile perioadei curente, este egal cu acelaşi indice, calculat cu ponderile perioadei de bază dacă cel puţin unul din factorii menţionaţi este egal cu zero;

2° — indicele de grup, calculat cu ponderile perioadei curente, este mai mare în valoare decît acelaşi indice, calculat cu ponderile perioadei de bază atunci cînd cei trei factori sînt diferiţi de zero şi coeficientul de corelaţie liniară simplă este pozitiv, deci ponderile se modifică în acelaşi sens cu valorile variabilei;

3° — indicele de grup, calculat cu ponderile perioadei curente, este mai mic decît cel calculat cu ponderile perioadei de bază atunci cînd cei trei factori sînt diferiţi de zero şi coeficientul de corelaţie liniară simplă are semn negativ, deci ponderile se modifică în sens invers cu valorile variabilei x.

Deci, aplicarea relaţiei lui Bortkiewicz la interpretarea indicilor statistici dă posibilitatea să se verifice gradul şi sensul în care sistemul de ponderare folosit influenţează asupra valorii indicilor.

11.3. DESCOMPUNEREA PE FACTORI A VARIAŢIEI UNUI FENOMEN COMPLEX FOLOSIND METODA INDICILOR

Problema descompunerii pe factori se pune atunci cînd se studiază variaţia unui fenomen complex ce depinde de variaţia mai multor factori. De exemplu, variaţia valorii producţiei se poate analiza în funcţie de volumul fizic şi de preţuri $V = f(q, p)$; a variaţiei volumului fizic al producţiei în funcţie de productivitatea muncii şi de cheltuiala totală de timp de muncă $q = f(w, T)$. Se pot întîlni cazuri cînd se înregistrează mai multe variabile-factori; de exemplu, cheltuielile totale de producţie cu materia primă (M) se analizează în funcţie de variaţia producţiei (q), a preţurilor de aprovizionare şi a normelor de consum (m), adică $M = f(m, p, q)$; sau productivitatea muncii medie lunară ($\bar{w}$) este în funcţie de productivitatea orară ($\bar{w}_h$), durata medie a zilei de lucru (d_z) şi durata medie a lunii de lucru (d_e), adică $\bar{w} = f(\bar{w}h, d_z, d_e)$.

Variaţia fenomenelor complexe se poate măsura atît în mărimi absolute (Δ), cît şi în mărimi relative ($\Delta I = R$).

În literatura de specialitate, descompunerea pe factori a sporului absolut şi relativ se mai numeşte şi descompunerea aritmetică sau analitică, iar descompunerea indicelui fenomenului complex în produsul indicilor factoriali, descompunere geometrică. Descompunerea analitică presupune o relaţie de însumare între factori, iar cea geometrică o relaţie de produs.

În teoria şi practica statistică, problemele care se pun cu prilejul descompunerii analitice şi geometrice se referă, în special, la analiza variaţiei unui fenomen complex luat în funcţie de două variabile-factori. Procedeele cele mai larg folosite sînt cele bazate pe „substituirea în lanţ" şi cel al

Scanned with OKEN Scanner

„repartizării proporționale a restului nedescompus" în funcție de gradul și sensul de influență ale factorilor. Pentru a nu ajunge uneori la rezultate contradictorii, se recomandă să se stabilească anumite analogii între descompunerea geometrică și cea analitică, respectiv să se folosească aceleași sisteme de ponderare.

Procedeul substituirii în lanț presupune să se anihileze pe rînd influența unui factor, păstrîndu-se numai variația factorului care interesează. Substituind în lanț factorii, înseamnă că pentru indicii factoriali se folosesc sisteme de ponderare diferite, iar numărul lor este egal cu cel al factorilor înregistrați.

Dacă, de pildă, se analizează un fenomen complex y în funcție de doi factori x și f și se folosește procedeul substituirii în lanț, înseamnă că se vor obține doi indici factoriali în care ponderea din perioada curentă folosită la unul din indici se înlocuiește cu ponderea perioadei de bază la cel de-al doilea indice. Indiferent care sînt variantele de ponderare, între cei trei indici există relația:

$$I^{y}_{1/0} = I^{y(x)}_{1/0} \cdot I^{y(f)}_{1/0}. \tag{11.47}$$

În mod corespunzător, la descompunerea analitică se stabilește relația:

$$\Delta^{y}_{1/0} = \Delta^{y(x)}_{1,0} + \Delta^{y(f)}_{1/0}. \tag{11.48}$$

Folosind sisteme de ponderare diferite pentru cei doi indici factoriali și, respectiv, pentru modificările absolute, înseamnă că se pot obține următoarele două variante:

Varianta I

$$I^{y(x)}_{1/0} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_1}; \quad \Delta^{y(x)}_{1/0} = \Sigma x_1 f_1 - \Sigma x_0 f_1; \tag{11.49}$$

$$I^{y(f)}_{1/0} = \frac{\Sigma x_0 f_1}{\Sigma x_0 f_0}; \quad \Delta^{y(x)}_{1/0} = \Sigma x_0 f_1 - \Sigma x_0 f_0. \tag{11.50}$$

Varianta II

$$I^{y(f)}_{1/0} = \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0}; \quad \Delta^{y(x)}_{1/0} = \Sigma x_1 f_0 - \Sigma x_0 f_0 \tag{11.51}$$

$$I^{y(f)}_{1/0} = \frac{\Sigma x_1 f_1}{\Sigma x_1 f_0}; \quad \Delta^{y(f)}_{1/0} = \Sigma x_1 f_1 - \Sigma x_1 f_0. \tag{11.52}$$

Pentru a interpreta în mod sugestiv legătura dintre variația fenomenului și variația celor doi factori se vor reprezenta grafic cele două variante luate la nivelul unei unități de observare — legătură care se păstrează și la nivelul întregului ansamblu.

Ambele variante prezentate presupun următoarele modificări ale fenomenului complex și ale factorilor:

$$y_1 > y_0; \; x_1 > x_0; \; f_1 > f_0,$$

Scanned with OKEN Scanner

pentru care se pot stabili următoarele sporuri:

— sporul fenomenului complex: $\Delta^{y}_{1/0} = y_1 - y_0$; (11.53)

— sporul factorului x: $\Delta^{x}_{1/0} = x_1 - x_0$; (11.54)

— sporul factorului f: $\Delta^{f}_{1/0} = f_1 - f_0$. (11.55)

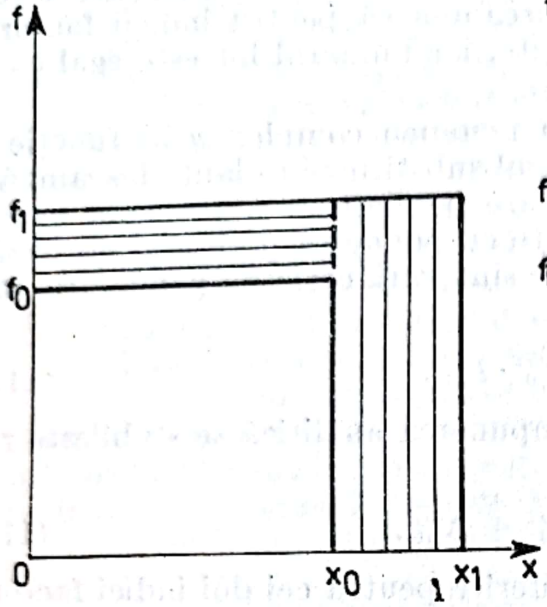


Fig. 11.1 Descompunerea pe factori folosind procedeul substituirii în lanț ($f_0 = f_1$; $x_1 = x_0$).

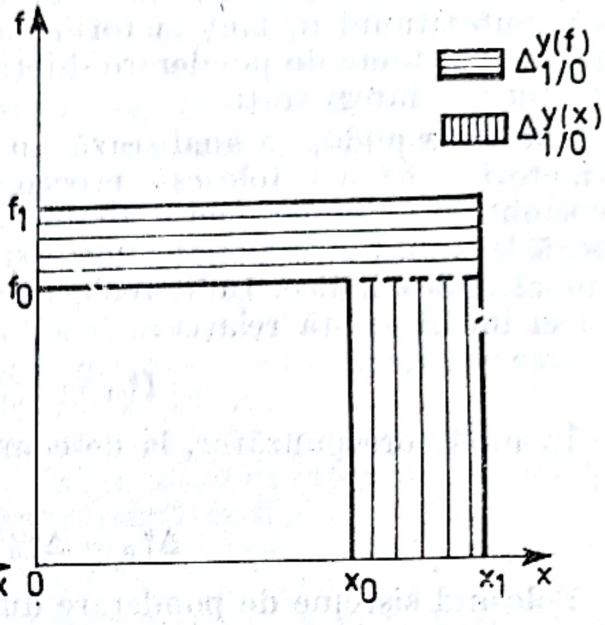


Fig. 11.2 Descompunerea pe factori folosind procedeul substituirii în lanț ($f_1 = f_0$; $x_0 = x_1$).

În varianta I, pentru măsurarea variației factorului x s-a folosit ponderea din perioada curentă, iar pentru variația factorului f ponderea din perioada de bază. Rezultă că se obțin următoarele modificări absolute:

— sporul variabilei y în funcție de x:

$$\Delta^{y(x)}_{1/0} = \Sigma x_1 f_1 - \Sigma x_0 f_1 = \Sigma[f_1(x_1 - x_0)] = \Sigma[f_1 \Delta^{x}_{1/0}]; \quad (11.56)$$

— sporul variabilei y în funcție de f:

$$\Delta^{y(f)}_{1/0} = \Sigma f_1 x_0 - \Sigma f_0 x_0 = \Sigma[x_0(f_1 - f_0)] = \Sigma[x_0 \Delta^{f}_{1/0}] \quad (11.57)$$

În varianta a II-a ponderile sînt folosite invers; pentru măsurarea variației lui x se folosește ponderea din perioada de bază, iar pentru variația factorului f se folosește ponderea din perioada curentă, adică:

— sporul variabilei y în funcție de x:

$$\Delta^{y(x)}_{1/0} = \Sigma x_1 f_0 - \Sigma x_0 f_0 = \Sigma[f_0(x_1 - x_0)] = \Sigma[f_0 \Delta^{x}_{1/0}]; \quad (11.58)$$

— sporul variabilei y în funcție de f:

$$\Delta^{y(f)}_{1/0} = \Sigma f_1 x_1 - \Sigma f_0 x_1 = \Sigma[x_1(f_1 - f_0)] = \Sigma(x_1 \Delta^{f}_{1/0}). \quad (11.59)$$

Dacă se are în vedere că $f_1 > f_0$, atunci înseamnă că în prima variantă sporul pe seama factorului x este mai mare decît în cea de-a doua variantă cu produsul diferenței dintre ponderi și creșterea variabilei x, adică

Scanned with OKEN Scanner

$\Delta^x(f_1 - f_0)$. Dacă se compară rezultatele celor două variante pentru cel de-al doilea factor, se constată că aici influenţa celui de-al doilea factor este mai mare cu aceeaşi valoare, deoarece ponderea folosită $x_1 > x_0$ se multiplică pentru toată creşterea factorului f, adică : $(x_1 - x_0)\Delta^f_{1/0}$. Cum $(x_1 - x_0)\Delta^f_{1/0} = (f_1 - f_0)\Delta^x$, înseamnă că, folosind procedeul substituirii în lanţ, o parte din sporul total al variabilei y se atribuie unuia din factori în raport cu mărimea şi sensul modificării factorului luat ca pondere. Cei mai mulţi dintre economişti şi statisticieni atribuie această parte influenţei factorului calitativ, folosind ca argument faptul că modificarea acestuia este condiţionată de modificarea factorului cantitativ şi deci se justifică folosirea ponderii din perioada curentă.

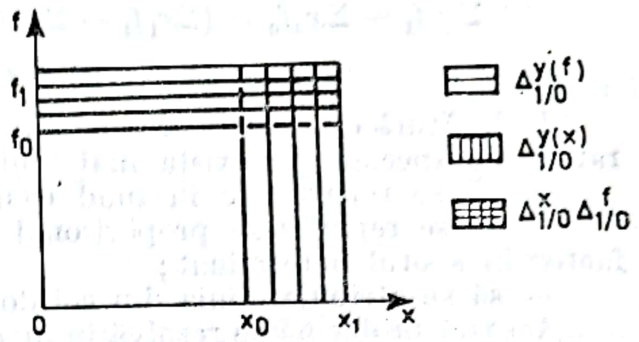


Fig. 11.3 Descompunerea pe factori folosind procedeul restului nedescompus.

Pornind de la acelaşi aspect, o seamă de economişti şi statisticieni consideră că, la descompunerea analitică şi geometrică, această parte să fie considerată ca rezultat al interacţiunii dintre cei doi factori. *Aceasta înseamnă să se determine iniţial, independent, influenţa celor doi factori folosind acelaşi sistem de ponderare pentru ambii factori (ponderarea din perioada de bază), iar apoi restul nedescompus să se repartizeze şi el pe cei doi factori de influenţă.* Pentru aceasta este necesar să se interpreteze graficul din figura 11.3.

Varianta a III-a presupune deci folosirea procedeului substituirii separate a factorilor. Relaţia determinantă pentru descompunerea geometrică este :

$$I^v_{1/0} = I^{v(x)}_{1/0} \cdot I^{v(f)}_{1/0} \cdot I^{v(x,f)}_{1/0}, \qquad (11.60)$$

în care :

$$\frac{\Sigma y_1}{\Sigma y_0} = \frac{\Sigma x_1 f_1}{\Sigma x_0 f_0}.$$

Se descompune astfel în produsul indicilor factoriali :

$$\frac{\Sigma x_1 f_1}{\Sigma f_0 x_0} = \left(\frac{\Sigma x_1 f_0}{\Sigma x_0 f_0}\right) \cdot \left(\frac{\Sigma f_1 x_0}{\Sigma f_0 x_0}\right) \cdot \left(\frac{\Sigma x_1 f_1}{\Sigma x_0 f_1} \cdot \frac{\Sigma x_1 f_0}{\Sigma x_0 f_0}\right) \cdot \qquad (11.61)$$

Comparînd cele trei formule de calcul ale indicilor factoriali se constată că primii doi sînt indici agregaţi care prezintă o singură influenţă, în timp ce al treilea este un indice format din raportul a doi indici agregaţi ai variabilei calitative cu ponderi diferite. Deci, al treilea indice poate fi considerat ca indice de tip Bortkiewicz, care reflectă gradul de influenţă a variaţiei ponderilor pentru modificarea de ansamblu a caracteristicii calitative cu care se găseşte în interacţiune.

Scanned with OKEN Scanner

De aici rezultă relaţia determinantă pentru descompunerea analitică:

$$\Delta^{y}_{1/0} = \Delta^{y(x)}_{1/0} + \Delta^{y(f)}_{1/0} + \Delta^{y(x,f)}_{1/0}, \tag{11.62}$$

în care:

$$\Delta^{y}_{1/0} = \Sigma y_1 - \Sigma y_0 = \Sigma x_1 f_1 - \Sigma x_0 f_0. \tag{11.63}$$

$$\Sigma x_1 f_1 - \Sigma x_0 f_0 = (\Sigma x_1 f_1 - \Sigma x_0 f_0) + (\Sigma x_0 f_1 - \Sigma x_0 f_0) + [(\Sigma x_1 f_1 - \Sigma x_0 f_1) - (\Sigma x_1 f_0 - \Sigma x_0 f_0)] \tag{11.64}$$

În legătură cu repartizarea pe factori a sporului nedescompus în literatura de specialitate există mai multe propuneri:

— să se repartizeze în mod egal pe factori;

— să se repartizeze proporţional cu gradul de influenţă a fiecărui factor în sporul determinat;

— să se atribuie unuia din cei doi factori de influenţă.

Această problemă se rezolvă în mod diferenţiat; de exemplu, în cazul comparărilor pe o perioadă scurtă, în mod frecvent se atribuie acest spor factorului calitativ, ceea ce conduce la folosirea procedeului substituirii în lanţ (varianta I). În cazul folosirii unei perioade mai mari se foloseşte frecvent varianta în care restul nedescompus se separă proporţional cu gradul de influenţă a celor doi factori.

În acest ultim caz se foloseşte o descompunere pe factori în două trepte:

— în prima treaptă se folosesc la calculul indicilor factoriali ponderi din perioada de bază pentru ambii factori şi restul nedescompus ($\Delta_x \Delta_y$) se consideră ca fiind rezultatul interacţiunii dintre factori (11.61) şi (11.62);

— în cea de-a doua treaptă se calculează partea care revine fiecărui factor din restul nedescompus şi se adaugă sporului iniţial.

Pentru a repartiza pe factori restul nedescompus, se calculează ponderea fiecărui factor (K^x, K^f) raportînd pe rînd sporul iniţial al factorilor ($\Delta y(x)$, $\Delta y(f)$) la suma lor care reprezintă partea repartizată.

Deci:

$$K^x = \frac{\Delta y(x)}{\Delta y(x) + \Delta y(f)} \text{ şi } K^f = \frac{\Delta y(f)}{\Delta y(x) + \Delta y(f)} \tag{11.65}$$

Sporul total al variabilei y în funcţie de factorul x şi în interdependenţă cu factorul f:

$$\Delta^{y(x,f)}_{total} = \Delta^{y(x)}_{iniţial} + K^x(\Delta^x \cdot \Delta^f)$$

$$\Delta^{y(x,f)}_{total} = \Sigma(x_1 f_0 - \Sigma x_0 f_0) + \frac{\Delta^{y(x)}}{\Delta^{y(x)} + \Delta^{y(f)}} \Delta^x \Delta^f. \tag{11.66}$$

Sporul total al variabilei în funcţie de factorul f şi în interdependenţă cu factorul x:

$$\Delta^{y(f,x)}_{total} = \Delta^{y(f)}_{iniţial} + K^f(\Delta^x \Delta^f) \tag{11.67}$$

$$\Delta^{y(f,x)}_{total} = \Sigma x_0 f_1 - \Sigma x_0 f_0 + \left(\frac{\Delta^{y(f)}}{\Delta^{y(x)} + \Delta^{y(f)}} \Delta^x \Delta^f \right) \tag{11.68}$$

Scanned with OKEN Scanner

Avînd calculată întreaga contribuţie a fiecărui factor, se poate calcula gradul de determinaţie a factorilor la sporul total al variabilei rezultative y.

Pentru a stabili proporţia în care cei doi factori influenţează modificarea absolută a variabilei complexe y, se calculează ponderea totală a fiecărui factor faţă de sporul total:

$$K^{x,f} = \frac{\Delta^{y(x,f)}}{\Delta^{y}} \cdot 100; \qquad (11.69)$$

$$K^{f,x} = \frac{\Delta^{y(f,x)}}{\Delta^{y}} \cdot 100, \qquad (11.70)$$

în care:

$K^{x,f}$ reprezintă coeficientul de determinaţie totală a factorului x în interdependenţă cu factorul f, folosind metoda indicilor;

$K^{f,x}$ — coeficientul de determinaţie totală a factorului f în interdependenţă cu factorul x asupra variabilei y, folosind metoda indicilor.

În cazul procedeului determinării influenţei separate a factorilor coeficientul de determinaţie totală este egal cu coeficientul ponderii determinat anterior.

Considerînd factorul x ca fiind de natură calitativă, se apreciază că dacă $K^{x,f} > K^{f,x}$, atunci calea de modificare a variabilei complexe este intensivă, iar pentru cazul în care $K^{x,f} < K^{f,x}$ calea de modificare este cea extensivă.

11.4. APLICAŢII LA CONSTRUIREA UNOR INDICI ECONOMICI

Aceste probleme teoretice se particularizează în funcţie de conţinutul indicatorului.

Pentru exemplificare se vor folosi indicii dinamicii producţiei şi productivităţii muncii.

11.4.1. Indicii valorii, volumului fizic şi ai preţurilor producţiei şi circulaţiei mărfurilor

În practica statistică, indicii valorii, volumului fizic şi ai preţurilor producţiei şi circulaţiei de mărfuri formează un sistem de indici, în cadrul căruia variaţia valorii este analizată în funcţie de modificarea volumului fizic şi a preţurilor. Acest sistem de indici poate să fie analizat în mod independent sau în interdependenţă cu alte sisteme de indici ce caracterizează alte laturi ale activităţii economice la nivelul unei întreprinderi, al unei ramuri de producţie sau la nivelul întregii economii naţionale.

Scanned with OKEN Scanner

Dacă se notează cantitatea de mărfuri cu q, prețul unitar cu p, iar valoarea cu v, rezultă următoarele relații:

— pentru perioada de bază:

$$v_0 = q_0 p_0; \tag{11.71}$$

— pentru perioada curentă:

$$v_1 = q_1 p_1. \tag{11.72}$$

Fiind trei elemente variabile, înseamnă că se pot construi trei indici, ce pot fi calculați ca indici individuali și ca indici de grup. Indicii individuali se calculează raportînd, la nivelul unităților observate, elementele din perioada curentă la cele din perioada de bază. Indicii de grup ai valorii, volumului fizic și ai prețurilor se calculează ca indici agregați.

Indicele individual al valorii se calculează raportînd valoarea din perioada curentă la valoarea din perioada de bază, folosind formula:

$$i^v_{1/0} = \frac{v_1}{v_0} = \frac{q_1 p_1}{q_0 p_0}. \tag{11.73}$$

Indicele de grup al valorii se calculează ca un indice agregat raportînd suma valorii mărfurilor din perioada curentă la suma valorii mărfurilor din perioada de bază, după formula:

$$I^v_{1/0} = \frac{\Sigma v_1}{\Sigma v_0} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_0}. \tag{11.74}$$

Calculul indicelui de grup al valorii mărfurilor nu ridică nici o problemă deosebită, deoarece valoarea este însumabilă direct pentru orice fel de producție, deci nu este necesar să se recurgă la un alt element care să joace rol de pondere. În plus, în toate formele de evidență există date cu privire la valoarea producției, respectiv a circulației de mărfuri, pentru orice perioadă de timp.

Pornind de la indicele de grup al valorii, se poate determina modificarea absolută de la o perioadă la alta, ca diferență între numărătorul și numitorul indicelui.

Modificarea absolută a unui fenomen (vezi capitolul „Analiza seriilor cronologice") se notează cu Δ și, în cazul valorii mărfii, se determină după formula:

$$\Delta^v_{1/0} = \Sigma q_1 p_1 - \Sigma q_0 p_0. \tag{11.75}$$

După semnul pe care îl are această diferență, ea poartă denumirea de spor (creștere) absolut(ă) a valorii producției sau de deficit (descreștere) absolut(ă) a valorii producției.

Indicele individual al volumului fizic arată de cîte ori s-a modificat, sub aspect cantitativ, producția din perioada curentă în raport cu cea din perioada de bază. Formula de calcul este:

$$i^q_{1/0} = \frac{q_1}{q_0} \tag{11.76}$$

Scanned with OKEN Scanner

Indicele de grup al volumului fizic se calculează tot ca un indice agregat şi trebuie să arate de cîte ori s-a modificat volumul fizic în perioada curentă faţă de perioada de bază pentru toată colectivitatea cercetată. Cum produsele se pot exprima în unităţi diferite de măsură, suma cantităţilor produse nu se poate face decît tot valoric. Pentru a putea surprinde numai modificarea de la o perioadă la alta a masei de mărfuri produse sau vîndute, este necesar — după cum s-a mai arătat — să se folosească aceleaşi preţuri şi la numărătorul şi la numitorul indicelui. Din punct de vedere matematic este suficient să se folosească în raport fie preţurile perioadei de bază, fie cele ale perioadei curente, adică să se folosească fie un indice de tip Laspeyres (11.77), fie un indice de tip Paasche (11.78).

Indicele volumului fizic calculat cu preţurile din perioada de bază:

$$I^q_{1/0} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}. \qquad (11.77)$$

Indicele volumului fizic calculat în preţurile din perioada curentă:

$$I^q_{1/0} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}. \qquad (11.78)$$

În mod firesc se pune problema: care dintre formule satisfac şi condiţiile impuse de conţinutul economic al indicatorului de comparat şi relaţiile obiective dintre elementele componente ale fenomenului complex a căror factori de influenţă sînt.

Analizînd conţinutul indicatorului şi relaţia care există între cantitate şi preţ se apreciază că indicele volumului fizic de tip Laspeyres (11.77) satisface în mai mare măsură aceste condiţii. De exemplu, dacă se calculează indicele dinamicii volumului fizic al producţiei anului 1977 faţă de producţia anului 1965 şi se folosesc ca ponderi preţurile perioadei de bază (1965), atunci se pune problema să se evalueze la aceste preţuri producţia anului curent. Dacă se folosesc ca ponderi preţurile curente, atunci se pune problema ca producţia anului de bază să se evalueze la preţurile anului curent. Or, este lipsită de sens economic evaluarea producţiei anului 1965 la preţurile anului 1977, atîta timp cît preţul reprezintă un element calitativ ce exprimă tocmai combinarea cheltuielilor de muncă vie cu cele de muncă materializată, în anumite condiţii de producţie date. Deci, folosind aceste preţuri s-ar include şi o parte din modificarea factorului calitativ. Este mai firesc, de aceea, să se considere că preţurile au rămas neschimbate şi că a avut loc o creştere cantitativă a producţiei.

În plus, acest indice prezintă avantajul că, cu ajutorul elementelor pe care le conţine, se poate determina corect şi mărimea modificărilor absolute realizate de la o perioadă la alta.

Modificarea absolută a volumului fizic al producţiei se notează cu $\Delta^q_{1/0}$ şi arată cu cît ar fi crescut (scăzut), în mărimi absolute, valoarea producţiei din perioada curentă faţă de cea din perioada de bază, dacă preţurile ar fi rămas aceleaşi. Deci, formula de calcul este:

$$\Delta^q_{1/0} = \Sigma q_1 p_0 - \Sigma q_0 p_0. \qquad (11.79)$$

Scanned with OKEN Scanner

În practica statistică, pentru determinarea indicelui volumului fizic se folosesc şi preţurile comparabile; indicele se va calcula după formula:

$$I^q_{1/0} = \frac{\Sigma q_1 p}{\Sigma q_0 p}, \tag{11.80}$$

în care:

p reprezintă preţul comparabil.

Indicele individual al preţului ($i^p_{1/0}$) se calculează raportînd preţul din perioada curentă la preţul din perioada de bază:

$$i^p_{1/0} = \frac{p_1}{p_0}. \tag{11.81}$$

Indicele preţurilor se calculează ca indice de grup ţinînd seama că preţul este un atribut al produselor respective. Deci, pe ansamblu trebuie avut în vedere şi cantităţile la care se referă structura producţiei vîndute. Elementul pondere este aici cantitatea din fiecare produs.

Şi aici apar două variante:

— indicele preţurilor de tip Laspeyres:

$$I^p_{1/0} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}; \tag{11.82}$$

— indicele preţurilor de tip Paasche:

$$I^p_{1/0} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}. \tag{11.83}$$

Din punct de vedere economic se consideră că indicele cu ponderile perioadei curente (11.83) satisface în mai mare măsură condiţiile de conţinut, deoarece variaţiile preţurilor sînt condiţionate de structura producţiei în care s-a făcut modificarea lor. În plus, acest indice dă posibilitatea măsurării corecte a volumului de economii sau a depăşirii care se obţin pe baza variaţiei preţurilor unitare. Dacă se are în vedere că în numărător este cuprinsă suma care a fost plătită pentru cumpărarea mărfurilor din perioada curentă, iar în numitor este valoarea care ar fi trebuit să fie plătită pentru aceeaşi cantitate dacă nu se modificau preţurile unitare, înseamnă că diferenţa dintre ele măsoară tocmai modificarea absolută a valorii produsă pe seama modificării preţurilor.

Economia (depăşirea) realizată pe seama modificării preţurilor ($\Delta^p_{1/0}$) se calculează după formula:

$$\Delta^p_{1/0} = \Sigma p_1 q_1 - \Sigma p_0 q_1. \tag{11.84}$$

Analizînd cei trei indici de grup, rezultă că, în practica statisticii din ţara noastră, indicele volumului fizic se calculează după formula propusă de Laspeyres, iar indicele preţurilor după formula propusă de Paasche.

Scanned with OKEN Scanner

Pe lîngă argumentele legate de conţinutul fenomenului de indexat, şi care au fost prezentate la interpretarea comparativă a celor doi indici, se mai adaugă şi unele aspecte legate de valoarea ponderilor utilizate.

Folosirea diferenţiată a celor două sisteme de ponderare satisface în plus şi relaţia de descompunere a indicelui valorii în produsul indicilor factoriali, care se verifică, în toate cazurile, pentru indicii individuali.

Deci, între cei trei indici există următoarea relaţie : indicele valorii este egal cu indicele volumului fizic înmulţit cu indicele preţurilor :

$$I^v_{1/0} = I^q_{1/0} \cdot I^p_{1/0}, \text{ de unde : } I^q_{1/0} = \frac{I^v_{1/0}}{I^p_{1/0}}, \text{ iar } I^p_{1/0} = \frac{I^v_{1/0}}{I^q_{1/0}}.$$

Dacă se înlocuiesc cei trei indici cu formulele lor de calcul se obţine :

$$\frac{\Sigma q_1 p_1}{\Sigma q_0 p_0} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \cdot \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0}.$$

Dacă cei doi indici factoriali vor fi calculaţi cu acelaşi sistem de ponderare, atunci modificarea valorii nu se mai descompune integral pe cei doi factori de influenţă şi apare un rest nedescompuns, care, în literatura de specialitate, se apreciază că este rezultatul interacţiunii dintre factori.

Această relaţie se foloseşte frecvent în practica statistică atunci cînd nu se pot calcula direct cei trei indici. Este suficient să se cunoască doi dintre cei trei indici pentru a se calcula cel de-al treilea.

Relaţia de produs a indicilor factoriali se transformă în relaţie de însumare a modificărilor absolute stabilite pe seama factorilor înregistraţi.

Modificarea absolută a valorii producţiei sau circulaţiei de mărfuri este egală cu suma modificărilor absolute calculate în funcţie de modificarea volumului fizic şi a preţurilor :

$$\Delta^v_{1/0} = \Delta^q_{1/0} + \Delta^p_{1/0}. \qquad (11.86)$$

Şi în acest caz este suficient să se cunoască două modificări absolute pentru a o determina pe cea de-a treia din cadrul sistemului de indici care a stat la baza calculării lor.

Modificarea absolută a volumului fizic este egală cu modificarea absolută a valorii, minus modificarea absolută a preţurilor :

$$\Delta^q_{1/0} = \Delta^v_{1/0} - \Delta^p_{1/0}. \qquad (11.87)$$

Modificarea absolută a preţurilor este egală cu modificarea absolută a valorii producţiei, minus modificarea absolută a volumului fizic al producţiei :

$$\Delta^p_{1/0} = \Delta^v_{1/0} - \Delta^q_{1/0}. \qquad (11.88)$$

Dar în practica statistică nu se dispune întotdeauna de toate elementele necesare calculării directe a indicilor agregaţi. Se pune problema să se înlocuiască indicii agregaţi prin indici sub formă de medie aritmetică sau armonică după datele cunoscute, astfel încît să îndeplinească condiţiile impuse de folosirea mediilor unor mărimi relative parţiale.

Scanned with OKEN Scanner

Atît pentru indicele de grup al volumului fizic, cît şi al preţurilor sub formă de agregat este necesar să se dispună la toate nivelurile de agregare de o evidenţă separată a cantităţilor vîndute. În practică, evidenţa cantităţilor vîndute se ţine doar pentru principalele produse, iar pentru majoritatea produselor numai în unităţi valorice. Dacă se are în vedere că şi la indicile de grup al volumului fizic, ca şi la cel al preţurilor trebuie să se recalculeze producţia dintr-o perioadă la preţul altei perioade, atunci este necesar să se folosească relaţiile dintre indicii individuali şi cei de grup astfel încît să se ajungă la acelaşi rezultat folosind cele două formule de calcul — şi sub formă de agregat şi sub formă de medie.

În cazul indicelui de grup trebuie să se respecte următoarele condiţii :

— să fie o medie a indicilor individuali ai volumului fizic, deci variabila x pentru care se calculează media este indicele individual al volumului fizic calculat la nivelurile de înregistrare folosite ;

— să fie egal cu indicele agregat al volumului fizic, deci trebuie ca numărătorul sau numitorul indicelui agregat să joace rol de ponderare.

Deci, în acest caz se dispune de următoarele relaţii derivate din formula indicelui agregat al volumului fizic :

$$\bar{x} = I^q_{1/0} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \; ; \quad x_i = i^q_{1/0} \; ; \quad f_i = q_0 p_0 \; ; \quad x_i f_i = q_1 p_0 .$$

Este evident faptul că pentru acest caz este avantajos să se folosească *indicele aritmetic ponderat al volumului fizic*, deoarece se cunoaşte valoarea din perioada de bază care constituie ponderea pentru această formă de medie :

$$I^q_{1/0} = \frac{\Sigma i^q_{1/0} q_0 p_0}{\Sigma q_0 p_0} . \tag{11.89}$$

Dacă în formula de mai sus se înlocuieşte indicele individual cu elementele sale de calcul se demonstrează că se ajunge la acelaşi rezultat ca şi în cazul cînd s-ar folosi un indice agregat :

$$I^q_{1/0} = \frac{\Sigma i^q_{1/0} q_0 p_0}{\Sigma q_0 p_0} = \frac{\Sigma \frac{q_1}{q_0} q_0 p_0}{\Sigma q_0 p_0} = \frac{\Sigma q_1 p_0}{q_0 p_0} .$$

Teoretic se poate stabili şi un indice al volumului fizic de forma unei medii armonice, dar care, practic, nu are sens, deoarece ar trebui să se ponderereze cu numărătorul indicelui ($q_1 p_0$) care necesită recalculări ale valorii mărfurilor din perioada curentă la preţurile perioadei de bază.

Indicele aritmetic al volumului fizic se calculează, deci, ca o medie aritmetică a indicilor individuali al acestuia ponderaţi cu valoarea din perioada de bază. El se foloseşte cînd nu se cunosc separat cantităţile şi preţurile, ci numai valoarea obţinută.

Pentru indicele preţurilor sub formă de medie se porneşte de la aceleaşi condiţii :

— să fie o medie a indicilor individuali ai preţurilor ;

— să fie o formă transformată a indicelui agregat al preţului.

Scanned with OKEN Scanner

Deci, în acest caz se dispune de următoarele date :

$$\bar{x} = I^p_{1/0} = \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0} \; ; \; x_i = i^p_{1/0} \; ; \; f_i = q_1 p_0 \; ; \; x_i f_i = q_1 p_1 .$$

Dacă se presupune cunoscută numai valoarea, înseamnă că pentru indicele preţurilor trebuie folosită o medie în care ponderarea se face cu produsele de frecvenţă xf şi care se găsesc în numărător — deci o medie de formă armonică. De asemenea, trebuie să se dispună de indicii individuali ai preţurilor proveniţi fie dintr-o observare totală, fie dintr-o cercetare selectivă.

Indicele armonic al preţurilor se calculează ca o medie armonică a indicilor individuali ai preţurilor ponderaţi cu valoarea mărfurilor vîndute în perioada curentă, adică :

$$I^p_{1/0} = \frac{\Sigma q_1 p_1}{\sum \frac{1}{i^p} q_1 p_1} \; . \qquad (11.90)$$

Şi în acest caz, dacă se înlocuieşte indicele individual cu formula sa de calcul, se ajunge la indicele agregat al preţurilor, adică :

$$I^p_{1/0} = \frac{\Sigma q_1 p_1}{\sum \frac{1}{i^p_{1/0}} q_1 p_1} = \frac{\Sigma q_1 p_1}{\sum \frac{1}{\frac{p_1}{p_0}} q_1 p_1} = \frac{\Sigma q_1 p_1}{\sum \frac{p_0}{p_1} q_1 p_1} = \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0} .$$

În concluzie, metodologia de calcul a celor trei indici care formează un sistem — în cadrul căruia se pot stabili relaţii bine determinate — este în funcţie, în primul rînd, de natura datelor de care se dispune.

Un prim caz — şi cel mai uşor de rezolvat — este acela în care se cunosc datele individuale cu privire la cantităţile şi preţurile mărfurilor vîndute. Această situaţie este specifică întreprinderilor industriale şi agricole care îşi desfac produsele către consumatorii productivi sau către comerţ. De asemenea, se poate întîlni această situaţie în circulaţia mărfurilor pentru produsele la care se ţine evidenţă şi în unităţile naturale.

Se presupune că într-o întreprindere se desfac trei produse, la care se dau următoarele date : (vezi tabelul 11.1).

Pe baza datelor din tabelul (11.1) se pot calcula direct indicii individuali şi de grup ai valorii, volumului fizic şi ai preţurilor :

— *indicii individuali ai valorii* :

$$i^v_{1/0} = \frac{q_1 p_1}{q_0 p_0} \begin{cases} \text{— A} \quad i^v_{1/0} = \dfrac{380\,000}{300\,000} \cdot 100 = 126{,}7\,\% \; ; \\ \text{— B} \quad i^v_{1/0} = \dfrac{648\,000}{600\,000} \cdot 100 = 108\,\% \; ; \\ \text{— C} \quad i^v_{1/0} = \dfrac{1\,300\,000}{1\,200\,000} \cdot 100 = 108{,}3\,\% \; ; \end{cases}$$

Scanned with OKEN Scanner

— *indicii individuali ai volumului fizic*:

$$i^q_{1/0} = \frac{q_1}{q_0} \begin{cases} \text{A} & i^q_{1/0} = \dfrac{2\,000}{1\,500} \cdot 100 = 135\%; \\ \text{B} & i^q_{1/0} = \dfrac{3\,600}{3\,000} \cdot 100 = 120\%; \\ \text{C} & i^q_{1/0} = \dfrac{10\,000}{8\,000} \cdot 100 = 125\%; \end{cases}$$

— *indicii individuali ai preţurilor*:

$$i^p_{1/0} = \frac{p_1}{p_0} \begin{cases} \text{A} & i^p_{1/0} = \dfrac{190}{200} \cdot 100 = 95\%; \\ \text{B} & i^p_{1/0} = \dfrac{180}{200} \cdot 100 = 90\%; \\ \text{C} & i^p_{1/0} = \dfrac{130}{150} \cdot 100 = 86{,}6\%. \end{cases}$$

Indicele de grup al valorii producţiei:

$$I^v_{1/0} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_0} \cdot 100 = \frac{2\,328\,000}{2\,100\,000} \cdot 100 = 110{,}7\%.$$

Creşterea absolută a valorii producţiei:

$$\Delta^v_{1/0} = \Sigma q_1 p_1 - \Sigma q_0 p_0 = 2\,328\,000 - 2\,100\,000 = +\,28\,000 \text{ lei.}$$

Indicele de grup al volumului fizic al producţiei:

$$I^q_{1/0} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \cdot 100 = \frac{2\,620\,000}{2\,100\,000} \cdot 100 = 124{,}7\%.$$

Creşterea absolută a volumului fizic al producţiei:

$$\Delta^q_{1/0} = \Sigma q_1 p_0 - \Sigma q_0 p_0 = 2\,620\,000 - 2\,100\,000 = +520\,000.$$

Indicele de grup al preţurilor:

$$I^p_{1/0} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \cdot 100 = \frac{2\,328\,000}{2\,620\,000} = 88{,}4\%.$$

Volumul economiilor realizate în urma reducerii preţurilor:

$$\Delta^p_{1/0} = \Sigma p_1 q_1 - \Sigma p_0 q_1 = 2\,328\,000 - 2\,620\,000 = -292\,000.$$

Scanned with OKEN Scanner

Tabelul 11.1

Cantitatea, preţul şi valoarea mărfurilor vîndute de întreprinderea *A* în perioada de bază şi curentă
(date convenţionale)

Denumirea produselor	Cantitatea vîndută (mii kg)		Preţul unitar (lei)		Valoarea mărfurilor vîndute (mii lei)		
	per. de bază	per. curentă	per. de bază	per. curentă	per. de bază	perioada curentă	
						la preţurile perioadei curente	la preţurile perioadei bază
	q_0	q_1	p_0	p_1	q_0p_0	q_1p_1	q_1p_0
A	1 500	2 000	200	190	300 000	380 000	400 000
B	3 000	3 600	200	180	600 000	648 000	720 000
C	8 000	10 000	150	130	1 200 000	1 300 000	1 500 000
	—	—	—	—	2 100 000	2 328 000	2 620 000
					Σq_0p_0	Σq_1p_1	Σq_1p_0

Scanned with OKEN Scanner

Deşi indicii individuali au arătat că dinamica cea mai rapidă a fost la primul produs, pe total valorile indicilor de grup se apropie de valoarea indicilor de la produsul C. Aceasta confirmă faptul că indicele de grup este o valoare medie în raport cu indicii individuali şi că oscilează ca valoare în jurul indicelui care are ponderea cea mai mare.

Pentru a verifica exactitatea calculelor se face produsul dintre indicele volumului fizic şi cel al preţurilor, care trebuie să fie aproximativ egal (din cauza cîturilor neexacte ale celor doi indici) cu indicele valoric, iar suma modificărilor absolute corespunzătoare celor doi indici factoriali trebuie să fie egală cu modificarea absolută a valorii :

$$1{,}247 \cdot 0{,}8847 = 1{,}10689 \simeq 1{,}107;$$

$$520\,000 + (-292\,000) = 228\,000 \text{ mii lei.}$$

Un al doilea caz ce se poate întîlni în practica statistică este acela în care se cunosc numai datele cu privire la valoarea mărfurilor şi dintr-o observare totală sau selectivă se cunosc indicii individuali ai volumului fizic.

Tabelul 11.2

Valoarea mărfurilor vîndute şi dinamica volumului fizic

	Valoarea mărfurilor vîndute (mii lei)		Indicii volumului fizic ai desfacerii mărfurilor
	perioada de bază	perioada curentă	
	q_0p_0	q_1p_1	$i^q_{1/0}$
1	1 200	1 300	1,10
2	2 300	2 500	1,07
3	2 000	2 000	1,12
Total	5 500	5 800	—
	Σq_0p_0	Σq_1p_1	

Pornind de la aceste date nu se poate calcula direct decît indicele de grup şi creşterea absolută a valorii circulaţiei de mărfuri, care vor fi egale cu :

$$I^v_{1/0} = \frac{\Sigma q_1p_1}{\Sigma q_0p_0} = \frac{5\,800}{5\,500} = 1{,}054 \text{ sau } 105{,}4\%;$$

$$\Delta^v_{1/0} = \Sigma q_1p_1 - \Sigma q_0p_0 = 5\,800 - 5\,500 = 300 \text{ mii lei.}$$

Indicele de grup al volumului fizic sub formă de agregat nu se poate calcula deoarece lipsesc cantităţile vîndute în perioada curentă. În schimb, cunoscînd indicii individuali ai volumului fizic şi valoarea circulaţiei măr-

Scanned with OKEN Scanner

furilor din perioada de bază se poate calcula indicele mediu aritmetic:

$$I^q_{i/0} = \frac{\Sigma i^q_{i/0}\, q_0 p_0}{\Sigma q_0 p_0} = \frac{(1{,}10 \cdot 1\,200) + (1{,}07 \cdot 2\,300) + (1{,}12 \cdot 2\,000)}{1\,200 + 2\,300 + 2\,000} =$$

$$= \frac{1\,320 + 2\,461 + 2\,240}{5\,500} = \frac{6\,021}{5\,500} = 1{,}0947 \text{ sau } 109{,}47\%.$$

Pornind de la formula de calcul al indicelui de grup se va putea determina creşterea absolută a volumului fizic al desfacerii în perioada curentă faţă de perioada de bază:

$$\Delta^q_{i/0} = \Sigma i^q_{i/0}\, q_0 p_0 - \Sigma q_0 p_0 = 6\,021 - 5\,500 = 521 \text{ mii lei.}$$

Pentru a determina indicele de grup al preţurilor se foloseşte relaţia care există între acest indice şi ceilalţi doi deja cunoscuţi:

$$I^p_{i/0} = \frac{I^v_{i/0}}{I^q_{i/0}} = \frac{1{,}054}{1{,}0947} = 0{,}963 \text{ sau } 96{,}3\%.$$

Economiile realizate pe seama modificării preţurilor se pot determina din relaţia acestui indicator cu celelalte două modificări absolute:

$$\Delta^p_{i/0} = \Delta^v_{i/0} - \Delta^q_{i/0} = 300 - 521 = -221 \text{ mii lei.}$$

Analizînd rezultatele obţinute se poate spune că dacă nu aveau loc reduceri de preţuri în perioada curentă faţă de perioada de bază, suma încasată din desfacerea mărfurilor ar fi crescut cu 9,47%, ceea ce corespunde unei creşteri absolute de 521 mii lei. În realitate, valoarea desfacerilor a crescut numai cu 5,4%, respectiv cu 300 mii lei, restul fiind tocmai economiile realizate de populaţie în urma reducerii preţurilor.

Un al treilea caz este acela în care s-a făcut o reducere de preţuri la mai multe grupe de mărfuri şi se pune problema să se calculeze procentul mediu de reducere şi suma economiilor realizate în urma modificării preţurilor.

Tabelul 11.3

Valoarea mărfurilor vîndute şi procentul de reducere a preţurilor

Grupe de mărfuri	Valoarea mărfurilor vîndute (mii lei)		% de reducere a preţurilor în perioada curentă faţă de perioada de bază
	perioada de bază	perioada curentă	
	$q_0 p_0$	$q_1 p_1$	$\left(I^p_{1/0} - 1\right)100$
A	100 000	110 000	− 5%
B	300 000	360 000	−10%
C	400 000	500 000	− 8%
Total	800 000	970 000	—
	$\Sigma q_0 p_0$	$\Sigma q_1 p_1$	

Scanned with OKEN Scanner

Procentul de reducere a prețurilor reprezintă tocmai ritmul scăderii, deci indicele se va obține adăugînd 100%.

Deci, pentru cele trei grupe de mărfuri se vor obține următorii indici parțiali ai prețurilor :

$$I^p_{1/0} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \begin{cases} - \text{A } i^p_{1/0} = 100\% + (-5\%) = 95\% \\ - \text{B } i^p_{1/0} = 100\% + (-10\%) = 90\% \\ - \text{C } i^p_{1/0} = 100\% + (-8\%) = 92\%. \end{cases}$$

Pe total indicele prețurilor nu se poate calcula direct sub formă de agregat, ci folosind indicele mediu armonic de forma :

$$I^p_{1/0} = \frac{\Sigma q_1 p_1}{\sum \frac{1}{i^p_{1/0}} q_1 p_1} = \frac{110\,000 + 360\,000 + 500\,000}{\left(\frac{1}{0{,}95} 110\,000\right) + \left(\frac{1}{0{,}90} 360\,000\right) + \left(\frac{1}{0{,}92} 500\,000\right)} =$$

$$= \frac{970\,000}{1\,058\,800} = 0{,}916 \text{ sau } 91{,}6\%.$$

Economiile realizate de populație în urma reducerii prețurilor :

$$\Delta^p_{1/0} = \Sigma q_1 p_1 - \sum \frac{1}{i^p_{1/0}} q_1 p_1 = 970\,000 - 1\,058\,800 = -88\,800 \text{ mii lei.}$$

Pe baza datelor din tabelul 11.3 se pot determina indicele valorii și creșterea absolută a valorii circulației de mărfuri :

$$I^v_{1/0} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_0} = \frac{970\,000}{800\,000} = 1{,}2125 \text{ sau } 121{,}25\% ;$$

$$\Delta^v_{1/0} = \Sigma q_1 p_1 - \Sigma q_0 p_0 = 970\,000 - 800\,000 = 170\,000 \text{ mii lei.}$$

Folosind relațiile dintre cei trei indici și, respectiv, cele trei modificări absolute, se pot determina indicele și creșterea absolută a volumului fizic :

$$I^q_{1/0} = \frac{I^v_{1/0}}{I^p_{1/0}} = \frac{1{,}2125}{0{,}916} = 1{,}323 \text{ sau } 132{,}3\% ;$$

$$\Delta^q_{1/0} = \Delta^v_{1/0} - \Delta^p_{1/0} = 170\,000 - (-88\,800) = 258\,800 \text{ mii lei.}$$

11.4.2. Indicii productivității muncii

Indicii raportului a două medii se calculează pentru a studia variația unei caracteristici derivate ce exprimă, cantitativ și calitativ, modul de formare a unei variabile statistice în funcție de valorile individuale sub

Scanned with OKEN Scanner

care se manifestă şi a frecvenţelor lor de apariţie. Printre indicii acestor caracteristici derivate, frecvent se întîlnesc : indicii productivităţii muncii, indicii costului mediu, indicii retribuţiei medii, indicii recoltei medii la ha etc.

Pornind de la formele generale ale indicelui raportului de medii (11.5) se trece la particularizarea lui în funcţie de conţinutul indicatorilor pentru care se calculează valoarea medie.

În cazul productivităţii muncii se porneşte de la modul de calcul al acestui indicator în funcţie de sfera de cuprindere şi de datele cuprinse în sistemul informaţional economic.

Nivelul productivităţii muncii se poate exprima direct prin cantitatea de produse ce revine în medie pe o unitate de timp sau, indirect, prin consumul de timp de muncă pe produs. În cel de-al doilea caz se mai poate defini şi ca fiind cantitatea de timp de muncă ce revine în medie pe o unitate de produs.

Rezultă de aici că, pentru a putea calcula nivelul şi dinamica productivităţii muncii, trebuie să se cunoască următoarele elemente :

q reprezintă cantitatea de produse obţinută în perioada pentru care se calculează productivitatea muncii ;

T — cheltuielile totale de timp de muncă consumat pentru producţia respectivă ;

w — nivelul productivităţii muncii exprimat prin cantitatea de produse obţinută într-o unitate de timp ;

t — consumul specific de muncă.

Nivelul productivităţii muncii w se calculează prin metoda directă raportînd volumul producţiei obţinute (q) la cheltuielile totale de timp de muncă (T), adică :

$$w = \frac{q}{T},$$

de unde :

$$q = wT.$$

Nivelul productivităţii muncii calculat prin metoda inversă (t) se obţine raportînd cheltuielile totale de timp de muncă la cantitatea totală de produse obţinute în perioada de timp la care se referă datele, adică :

$$t = \frac{T}{q}, \text{ în care } T = tq.$$

Dacă în primul raport se înlocuieşte T cu echivalentul său, în funcţie de t, se obţine :

$$w = \frac{q}{T} = \frac{q}{tq} = \frac{1}{t}.$$

Deci, nivelul productivităţii muncii prin metoda directă (w) este invers proporţional cu nivelul consumului specific de timp de muncă (t).

Scanned with OKEN Scanner

De aici rezultă că, pentru măsurarea variației în timp și spațiu a productivității muncii, se pot folosi două sisteme de indici.

În cazul exprimării directe a productivității muncii se vor obține indici individuali și indici de grup :

— indicele individual al productivității muncii :

$$i^{w}_{1/0} = \frac{w_1}{w_0} \; ; \tag{11.91}$$

— indicele de grup se va calcula raportînd nivelul mediu al productivității muncii din perioada curentă la cel din perioada de bază :

$$I^{\bar{w}}_{1/0} = \frac{\bar{w}_1}{\bar{w}_0} \; , \tag{11.92}$$

în care :

$\bar{w}$ reprezintă media aritmetică ponderată a nivelurilor individuale ale productivității muncii ponderate cu cheltuielile de timp de muncă, adică variabila $x = w$, iar frecvențele valorilor $f = T$.

În formă explicită, indicele de grup al productivității muncii se poate scrie :

$$I^{\bar{w}}_{1/0} = \frac{\Sigma w_1 T_1}{\Sigma T_1} : \frac{\Sigma w_0 T_0}{\Sigma T_0} \; . \tag{11.93}$$

Rezultă de aici că indicele de grup al productivității muncii este un indice cu structură variabilă care arată că variația lui $\bar{w}_0 \rightarrow \bar{w}_1$ depinde de modificarea valorilor individuale $w_0 \rightarrow w_1$ și de modificarea frecvențelor de apariție $T_0 \rightarrow T_1$.

Dacă, de exemplu, se analizează variația productivității muncii într-o ramură în funcție de nivelurile de productivitate a muncii de la fiecare întreprindere, atunci trebuie să se țină seama și de cantitatea totală de timp de muncă care s-a depus în fiecare întreprindere. Aceasta se poate exprima prin ore—om, zile—om sau număr mediu sau efectiv al muncitorilor sau al întregului personal muncitor calculat pe o perioadă (lună, trimestru, semestru, an).

Dacă se are în vedere faptul că această colectivitate a muncitorilor este foarte dinamică, atît ca volum, cît și ca mutații de structură (vechime, calificare, înzestrare tehnică etc.) înseamnă că asupra nivelului mediu al productivității muncii influențează, pe lîngă variația valorilor individuale calculate la nivelul unităților complexe înregistrate, și deplasările de structură produse prin modificarea frecvențelor de la o grupă la alta.

Este suficient să se amintească aici două tendințe care se întîlnesc frecvent în industrie : realizarea procesului de creștere continuă a calificării, care se concretizează în trecerea muncitorilor către o grupă superioară de calificare și procesul obiectiv al reînnoirii forței de muncă prin intrarea în producție a noilor contingente de muncitori și care, de regulă, pentru perioada de adaptare, măresc ponderea celor cu un nivel mai scăzut de productivitate. Deci, deplasările de structură se pot produce în ambele

Scanned with OKEN Scanner

sensuri. După cum se face compensarea tendinţelor pe ansamblu, modificarea de structură poate influenţa în sensul creşterii valorii medii sau al reducerii ei. De aceea este necesar să se analizeze separat influenţa factorului calitativ — productivitatea individuală a muncii — de influenţa factorului cantitativ — cheltuielile totale de timp de muncă pentru fiecare productivitate individuală a muncii.

Influenţa factorului calitativ se calculează folosind un *indice cu structură fixă al productivităţii muncii*. După cum s-a arătat, din punct de vedere matematic acesta se poate calcula folosind structura perioadei de bază sau pe cea din perioada curentă. Din punct de vedere economic trebuie analizat conţinutul indicatorului şi cu scopul pentru care se calculează. Dacă interesează colectivitatea actuală şi modul în care aceasta influenţează asupra modificării volumului producţiei, atunci se foloseşte structura cheltuielilor de timp din perioada curentă. După datele de care se dispune, indicele cu structură fixă al productivităţii muncii se poate calcula prin următoarele trei variante :

$$I_{1/0}^{\bar{w}(w)} = \frac{\Sigma w_1 T_1}{\Sigma T_1} : \frac{\Sigma w_0 T_1}{\Sigma T_1} = \frac{\overset{(1)}{\Sigma w_1 T_1}}{\Sigma w_0 T_1} = \frac{\overset{(2)}{\Sigma w_1 T_1}}{\Sigma \frac{1}{i^w} w_1 T_1} = \frac{\overset{(3)}{\Sigma w_1 T_1}}{\Sigma \frac{1}{i^w} w_1 T_1}. \qquad (11.94)$$

Alegerea primei formule, păstrînd forma de raport a două medii, se face atunci cînd interesează să se calculeze care ar fi fost nivelul mediu al productivităţii muncii pe total dacă muncitorii încadraţi în perioada curentă lucrau cu productivitatea muncii din perioada de bază.

Cea de-a doua formulă, folosind un indice agregat al producţiei calculat numai sub influenţa productivităţii muncii, se foloseşte atunci cînd cei trei indici ai productivităţii muncii sînt consideraţi ca un subsistem ai sistemului de indici mai cuprinzător ai producţiei de mărfuri.

Pe baza acestui indice se poate calcula şi modificarea volumului fizic al producţiei ca urmare a creşterii productivităţii muncii :

$$\Delta_{1/0}^{q(w)} = \Sigma w_1 T_1 - \Sigma w_0 T_1. \qquad (11.95)$$

Această diferenţă exprimă relaţia dintre volumul producţiei efective din perioada curentă ($w_1 T_1$) şi estimarea producţiei din perioada curentă, dacă muncitorii în această perioadă ar fi lucrat cu productivitatea muncii din perioada de bază ($w_0 T_1$).

Cel de-al treilea indice se foloseşte cînd se cunosc indicii individuali ai productivităţii muncii şi se cere indicele pentru întreaga colectivitate. Pornind de la indicele agregat este evident că dintre cele două şiruri de valori cuprinse în numărător ($w_1 T_1$) şi în numitor ($w_0 T_1$) se cunosc numai cele din numărător, ceea ce, de fapt, reprezintă volumul fizic din perioada curentă, deci trebuie aplicată media armonică.

Pentru calcularea influenţei factorului cantitativ se foloseşte indicele de variaţie a structurii colectivităţii cheltuielilor totale de timp de muncă, după formula :

$$I_{1/0}^{\bar{w}(T)} = \frac{\Sigma w_0 T_1}{\Sigma T_1} : \frac{\Sigma w_0 T_0}{\Sigma T_0}. \qquad (11.96)$$

Scanned with OKEN Scanner

Între cei trei indici există relaţia : indicele cu structură variabilă este egal cu produsul indicilor cu structură fixă şi cel de variaţie a structurii :

$$I_{1/0}^{\bar{w}} = I_{1/0}^{\bar{w}(w)} \cdot I_{1/0}^{\bar{w}(T)}, \tag{11.97}$$

de unde :

$$I_{1/0}^{\bar{w}(w)} = \frac{I_{1/0}^{\bar{w}}}{I_{1/0}^{\bar{w}(T)}}$$

şi respectiv :

$$I_{1/0}^{\bar{w}(T)} = \frac{I_{1/0}^{\bar{w}}}{I_{1/0}^{\bar{w}(w)}}.$$

Deci şi în acest caz este suficient să se cunoască numai doi indici ca să se determine valoarea celui de-al treilea.

În mod analog se poate face şi descompunerea modificărilor absolute pornind de la cei trei indici calculaţi sub formă de raport de medii şi făcînd diferenţele dintre prima şi a doua fracţie.

Acestea sînt formulele generale de calcul ale indicilor productivităţii muncii. La aplicarea lor apar însă probleme legate de modul de exprimare a indicatorilor primari din care se calculează productivitatea muncii, respectiv volumul producţiei şi cheltuielile totale de timp de muncă $\left(w = \frac{q}{T}\right)$.

În funcţie de modul de exprimare al producţiei, productivitatea muncii se poate calcula după metoda :

— unităţilor naturale ;
— unităţilor natural-convenţionale ;
— unităţilor valorice ;
— unităţilor de timp de muncă.

Metoda unităţilor naturale se foloseşte cînd producţia este omogenă şi, deci, însumabilă direct la orice nivel de înregistrare. În acest caz formula de calcul este :

— indicele individual :

$$i_{1/0}^{w} = \frac{w_1}{w_0} = \frac{q_1}{T_1} : \frac{q_0}{T_0}; \tag{11.98}$$

— indicele de grup :

$$I_{1/0}^{\bar{w}} = \frac{\bar{w}_1}{\bar{w}_0} = \frac{\Sigma w_1 T_1}{\Sigma T_1} : \frac{\Sigma w_0 T_0}{\Sigma T_0} = \frac{\Sigma q_1}{\Sigma T_1} : \frac{\Sigma q_0}{\Sigma T_0}. \tag{11.99}$$

Deci, la această metodă există relaţiile pe fiecare nivel de înregistrare :

$q_0 = w_0 T_0$ şi $q_1 = w_1 T_1$.

Metoda unităţilor natural-convenţionale se foloseşte atunci cînd producţia este asemănătoare din punctul de vedere al materiei prime, al procesului tehnologic şi al destinaţiei produselor. Aceasta înseamnă că pro-

ducția nu este însumabilă direct, dar admite un coeficient de echivalență (K), ceea ce presupune că înainte de a calcula productivitatea muncii se calculează producția diferită în unități convenționale. Formulele de calcul ale indicilor vor fi :

— indicele individual :

$$i^{w}_{1/0} = \frac{w_1}{w_0} = \frac{q_1 K}{T_1} : \frac{q_0 K}{T_0}; \qquad (11.100)$$

— indicele de grup :

$$I^{\overline{w}}_{1/0} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{\Sigma w_1 T_1}{\Sigma T_1} : \frac{\Sigma w_0 T_0}{\Sigma T_0} = \frac{\Sigma q_1 K}{\Sigma T_1} : \frac{\Sigma q_0 K}{\Sigma T_0}. \qquad (11.101)$$

La această metodă există relațiile corespunzătoare pentru fiecare nivel de înregistrare între producție, productivitate și cheltuielile totale de timp de muncă, $q_0 K = w_0 T_0$ și $q_1 K = w_1 T_1$.

Metoda unităților valorice se folosește pentru producția eterogenă și se bazează pe exprimarea producției obținute în prețuri comparabile :

— indicele individual :

$$i^{w}_{1/0} = \frac{w_1}{w_0} = \frac{q_1 p}{T_1} : \frac{q_0 p}{T_0}; \qquad (11.102)$$

— indicele de grup :

$$I^{\overline{w}}_{1/0} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{\Sigma w_1 T_1}{\Sigma T_1} : \frac{\Sigma w_0 T_0}{\Sigma T_0} = \frac{\Sigma q_1 p}{\Sigma T_1} : \frac{\Sigma q_0 p}{\Sigma T_0}. \qquad (11.103)$$

De aici se pot stabili relațiile : $q_0 p = w_0 T_0$ și $q_1 p = w_1 T_1$.

În practica statistică metoda unităților valorice este metoda oficială de calcul al nivelului și dinamicii productivității muncii și se aplică la nivelul fiecărei întreprinderi în funcție de producția netă industrială și de fondul de timp de muncă al întreprinderii.

Metoda unităților de timp de muncă, din punct de vedere teoretic, se poate aplica pentru orice fel de producție. Practic, această metodă nu se poate aplica decît dacă există evidență asupra normei de consum specific de timp de muncă pentru fiecare produs (t).

Documentele Plenarei C.C. al P.C.R. din 22—23 martie 1978 prevăd introducerea obligatorie a evidenței consumului de timp de muncă pe unitate de produs, care va sta la baza calculării nivelului de productivitate a muncii în întreprinderi.

Formulele de calcul sînt :

— indicele individual :

$$i^{w}_{1/0} = \frac{w_1}{w_0} = \frac{q_1 t_n}{T_1} : \frac{q_0 t_n}{T_0}; \qquad (11.104)$$

Scanned with OKEN Scanner

— indicele de grup :

$$I_{1/0}^{\overline{w}} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{\Sigma w_1 T_1}{\Sigma T_1} : \frac{\Sigma w_0 T_0}{\Sigma T_0} = \frac{\Sigma q_1 t_n}{\Sigma T_1} : \frac{\Sigma q_0 t_n}{\Sigma T_0}. \qquad (11.105)$$

La această metodă există relaţiile :

$$q_0 t_n = w_0 T_0 ; \quad q_1 t_n = w_1 T_1.$$

Pentru indicii de grup cu structură fixă şi de variaţie a structurii se folosesc formulele generale în care nivelurile individuale ale productivităţii se vor înlocui cu formulele respective aplicate în funcţie de gradul de omogenitate al producţiei.

Indicii productivităţii muncii se clasifică şi în funcţie de modul de exprimare a cheltuielilor de timp de muncă. Din acest punct de vedere ei pot fi : ai productivităţii muncii orare, zilnice, lunare, trimestriale, semestriale şi anuale. Toţi aceşti indici pot fi incluşi într-un sistem de indici în care indicele productivităţii lunare poate fi luat în funcţie de variaţia productivităţii orare şi a gradului de folosire a zilei şi lunii de lucru. Pe acelaşi principiu se poate analiza şi variaţia productivităţii muncii anuale în funcţie de factorii intensivi şi extensivi care o determină.

În continuare se vor lua cîteva cazuri de aplicare a indicilor productivităţii muncii pentru o producţie omogenă şi pentru o producţie eterogenă.

În cazul producţiei omogene se presupune că există un produs care se fabrică la două întreprinderi şi la care s-au înregistrat datele necesare calculării nivelului şi dinamicii productivităţii muncii pe fiecare întreprindere şi pe total.

Tabelul 11.4

Calculul indicilor productivităţii muncii în unităţi naturale

Întreprinderea	Producţia obţinută (bucăţi)		Nr. mediu		Nivelul productivităţii muncii		Indicii prod. muncii	Estimarea producţiei crt. pe baza prod. muncii din per. de bază
	per. de bază	per. curentă	per. de bază	per. curentă	per. de bază	per. curentă		
	q_0	q_1	T_0	T_1	$w_0 = \frac{q_0}{T_0}$	$w_1 = \frac{q_1}{T_1}$	$iw = \frac{w_1}{w_0}$	$w_0 T_1$
0	1	2	3	4	5	6	7	8
I	240 000	192 000	600	400	400	480	1,20	160 000
II	240 000	396 000	400	600	600	660	1,10	360 000
	480 000 $\Sigma q_0 = \Sigma w_0 T_0$	588 000 $\Sigma q_1 = \Sigma w_1 T_1$	1 000 ΣT_0	1 000 ΣT_1	480 $\overline{w}_0$	588 $\overline{w}_1$	1,225 $I_{1/0}^{\overline{w}}$	520 000 $\Sigma w_0 T_1$

Scanned with OKEN Scanner

Indicii individuali ai productivității muncii (col. 7) s-au calculat astfel :

$$i^{w}_{1/0} = \frac{w_1}{w_0} \begin{cases} \text{I} & i^{w}_{1/0} = \dfrac{480}{400} = 1{,}20 \text{ sau } 120\%, \\ \text{II} & i^{w}_{1/0} = \dfrac{660}{600} = 1{,}10 \text{ sau } 110\%, \end{cases}$$

iar indicele de grup după formula :

$$I^{w}_{1/0} = \frac{\Sigma\, q_1}{\Sigma\, T_1} : \frac{\Sigma\, q_0}{\Sigma\, T_0} = \frac{588\,000}{1\,000} : \frac{480\,000}{1\,000} = \frac{588}{480} = 1{,}225 \text{ sau } 122{,}5\%.$$

Analizînd rezultatele obținute se poate vedea că, deși la prima întreprindere productivitatea muncii a crescut cu 20% și la a doua întreprindere cu 10%, totuși pe ansamblu creșterea productivității muncii devansează creșterile la ambele întreprinderi. Aceasta este așa-numitul „*paradox statistic*" în care indicele de grup poate să se plaseze în afara intervalului de variație a valorilor indicilor individuali. Acest lucru atestă faptul că indicele de grup al productivității muncii este un indice cu structură variabilă, care, pe lîngă influența creșterii productivității individuale a muncii, cuprinde și influența unui alt factor, care este tocmai variația de structură.

Pentru a stabili gradul de influență a celor doi factori se vor calcula și ceilalți doi indici de grup.

Indicele productivității muncii cu structură fixă :

$$I^{\bar{w}(w)}_{1/0} = \frac{\Sigma\, w_1 T_1}{\Sigma\, T_1} : \frac{\Sigma\, w_0 T_1}{\Sigma\, T_1} = \frac{588\,000}{1\,000} : \frac{520\,000}{1\,000} = \frac{588}{520} = 1{,}1307 \text{ sau } 113{,}07\%.$$

Dacă nu s-ar fi produs variații în structura colectivității de muncitori, atunci, pe baza creșterii productivității muncii de la fiecare întreprindere, nivelul mediu pe total ar fi crescut cu 13,07%. De aici se poate trage concluzia că și cel de-al doilea factor are o influență pozitivă și va avea o valoare supraunitară :

$$I^{w(T)}_{1/0} = \frac{\Sigma\, w_0 T_1}{\Sigma\, T_0} : \frac{\Sigma\, W_0 T_0}{\Sigma\, T_0} = \frac{520\,000}{1\,000} : \frac{480\,000}{1\,000} = \frac{520}{480} = 1{,}0833 \text{ sau } 108{,}33\%.$$

În cifre absolute productivitatea muncii a crescut astfel :

— sub influența ambilor factori :

$$I^{\bar{w}}_{1/0} = \bar{w}_1 - \bar{w}_0 = \frac{\Sigma\, w_1 T_1}{\Sigma\, T_1} - \frac{\Sigma\, w_0 T_0}{\Sigma\, T_0} = 588 - 480 = +10 \text{ buc./muncitor ;}$$

— sub influenţa productivităţii muncii de la fiecare întreprindere :

$$I_{1/0}^{\bar{w}(w)} = \frac{\Sigma w_1 T_1}{\Sigma T_1} - \frac{\Sigma w_0 T_1}{\Sigma T_1} = 588 - 520 = +68 \text{ buc./muncitor};$$

— ca urmare a creşterii ponderii muncitorilor cu o productivitate mai ridicată de la 40 % $\left(\frac{400}{1\,000}\right)$ în perioada de bază la 60 % $\left(\frac{600}{1\,000}\right)$ în perioada curentă :

$$I_{1/0}^{w(T)} = \frac{\Sigma w_0 T_1}{\Sigma T_1} - \frac{\Sigma w_0 T_0}{\Sigma T_0} = 520 - 480 = +40 \text{ buc./muncitor.}$$

Creşterea producţiei pe seama creşterii productivităţii muncii va fi :

$$\Delta_{1/0}^{q(w)} = \Sigma w_1 T_1 - \Sigma w_0 T_1 = 588\,000 - 520\,000 = +68\,000 \text{ bucăţi.}$$

Acest caz a fost cel mai simplu, în care producţia fiind omogenă, era însumabilă direct.

Cel de-al doilea caz — producţia eterogenă — se rezolvă în funcţie de elementul folosit pentru omogenizare : coeficient de echivalenţă, preţuri comparabile sau cheltuieli de timp de muncă. Faţă de primul caz, este necesar de găsit mijlocul de a face însumabilă producţia, celelalte operaţii fiind aceleaşi. Pentru exemplificare se va lua numai calculul productivităţii muncii în unităţi natural-convenţionale. Deci, în acest caz, producţia în unităţi natural-fizice se va înmulţi cu coeficientul de transformare şi se va obţine volumul producţiei în unităţi natural-convenţionale, care, raportat la cheltuielile totale de timp de muncă, va da nivelul productivităţii muncii (vezi tabelul 11.5).

Analizînd rezultatele obţinute, se constată că dinamica cea mai rapidă a productivităţii muncii s-a înregistrat la produsul B (141,16 %) urmată de produsul C(125 %) şi apoi la produsul A(120 %).
Aceşti indici cu valori destul de ridicate se regăsesc în dinamica productivităţii muncii pe total (132,36 %).

Pentru a stabili influenţa celor doi factori se calculează şi aici indicele cu structură fixă şi cel de variaţie a structurii. Pentru aceasta este necesar să se estimeze volumul producţiei în unităţi convenţionale din perioada curentă, dacă muncitorii din această perioadă lucrau cu productivitatea perioadei de bază ($w_0 T_1$).

Indicele productivităţii muncii cu structură fixă :

$$I_{1/0}^{\bar{w}(w)} = \frac{\Sigma w_1 T_1}{\Sigma T_1} : \frac{\Sigma w_0 T_1}{\Sigma T_1} = \frac{22\,530}{1\,100} : \frac{18\,250}{1\,100} = \frac{21{,}84}{16{,}6} = 1{,}313.$$

Scanned with OKEN Scanner

Tabelul 11.5

Calcularea indicilor productivității muncii în unități de timp de muncă

Produse	Coeficient de echivalență	Producția obținută (mii buc.)		Număr mediu al muncitorilor		Producția în unități natural-convenționale (mii)		Nivelul productivității muncii		Indicii productivității muncii	Estimarea producției per. crt. în funcție de productivitatea muncii din perioada de bază
		per. de bază	per. curentă	per. de bază	per. curentă	per. de bază	per. curentă	per. de bază	per. curentă		
	K	q_0	q_1	T_0	T_1	q_0K	q_1K	$w_0 = \frac{q_0K}{T_0}$	$w_1 = \frac{q_1K}{T_1}$	$i^w_{1/0} = \frac{w_1}{w_0}$	w_0T_1
0	1	2	3	4	5	6	7	8	9	10	11
A	1,0	3 000	3 700	200	210	3 000	3 780	15	18	1,20	3 150
B	1,5	5 000	8 000	300	340	7 500	12 000	25	35,29	1 4116	8 500
C	2,0	3 000	4 125	500	550	6 000	8 250	12	15	1,25	6 600
Total	—	—	—	1 000 ΣT_0	1 100 ΣT_1	16 500 $\Sigma q_0k = \Sigma w_0T_0$	24 030 $\Sigma q_1k = \Sigma w_1T_1$	16,50 $\bar{w}_0$	21,84 $\bar{w}_0$	1,3236 $I^{\bar{w}}_{1/0}$	18 250 Σw_0T_1

Pe baza datelor din tabel s-au calculat nivelurile productivității muncii din perioada de bază și curentă raportînd pe rînd valorile din col. 6 și 7 la valorile din col. 4 și 5, precum și indicii individuali și indicele de grup cu structură variabilă, adică:

$$I^{\bar{w}}_{1/0} = \frac{\bar{w}_1}{w_0} = \frac{\Sigma w_1 T_1}{\Sigma T_1} : \frac{\Sigma w_0 T_0}{\Sigma T_0} = \frac{\Sigma q_1 K}{\Sigma T_1} : \frac{\Sigma q_0 K}{\Sigma T_0}; \quad i^w_{1/0} = \frac{w_1}{w_0} = \frac{q_0k}{T_1} : \frac{q_0k}{T_0}$$

A $i^w_{1/0} = \frac{18}{15} = 1,20$

B $i^w_{1/0} = \frac{35,29}{25} = 1,4116$

C $i^w_{1/0} = \frac{15}{12} = 1,25$

$$I^{\bar{w}}_{1/0} = \frac{24\,030}{1\,100} : \frac{16\,500}{1\,000} = \frac{21,84}{16,50} = 1,3236$$

Scanned with OKEN Scanner

Pentru verificare, indicele cu structură fixă se poate calcula şi ca medie armonică, după formula :

$$I_{1/0}^{\overline{ic}(w)} = \frac{\Sigma w_1 T_1}{\Sigma \frac{1}{i_{1/0}^{w}} w_1 T_1} = \frac{3\,780 + 12\,000 + 8\,250}{\frac{1}{1{,}20} 3\,780 + \frac{1}{1{,}4116} 12\,000 + \frac{1}{1{,}25} 8\,250} =$$

$$= \frac{24\,030}{3\,150 + 8\,500{,}9 + 6\,600} = \frac{24\,030}{18\,250{,}9} = 1{,}1313.$$

Indicele productivităţii muncii a modificării structurale :

$$I_{1/0}^{\overline{ic}(T)} = \frac{\Sigma w_0 T_1}{\Sigma T_1} : \frac{\Sigma w_0 T_0}{\Sigma T_0} = \frac{18\,250}{1\,100} : \frac{16\,500}{1\,000} = \frac{16{,}6}{16{,}5} = 1{,}006.$$

Indicii productivităţii muncii se mai pot calcula şi prin metoda inversă, pe baza relaţiei care există între consumul specific de timp de muncă şi nivelul productivităţii muncii. Această relaţie de invers proporţionalitate se păstrează în cazul indicilor individuali şi de grup.

Indicele individual al productivităţii muncii este invers proporţional cu indicele consumului specific de timp de muncă, adică :

$$i_{1/0}^{w} = \frac{1}{i_{1/0}^{t}} = \frac{1}{\frac{t_1}{t_0}}, \text{ de unde : } i_{1/0}^{w} = \frac{t_0}{t_1}. \qquad (11.106)$$

Rezultă că, pentru ca să crească nivelul productivităţii muncii, trebuie să scadă cheltuiala de timp de muncă pe unitate de produs.

Indicele de grup al productivităţii muncii este şi el invers proporţional cu indicele cheltuielilor medii de timp de muncă, adică :

$$I_{1/0}^{\overline{w}} = \frac{1}{I_{1/0}} = \frac{1}{\frac{\bar{t}_1}{\bar{t}_0}} = \frac{\bar{t}_0}{\bar{t}_1}, \qquad (11.107)$$

în care :

$$\bar{t}_0 = \frac{\Sigma t_0 q_0}{\Sigma q_0} \text{ iar } \bar{t}_1 = \frac{\Sigma t_1 q_1}{\Sigma q_1}.$$

Indicele de grup al productivităţii muncii va avea formula :

$$I_{1/0}^{\overline{w}} = \frac{\Sigma t_0 q_0}{\Sigma q_0} : \frac{\Sigma t_1 q_1}{\Sigma q_1}. \qquad (11.108)$$

Scanned with OKEN Scanner

În acest caz factorul cantitativ se referă la structura producţiei (q), deci indicele cu structură variabilă va reflecta atît influenţa modificării $t_0 \to t_1$, cît şi frecvenţele cu care se calculează media ($q_0 \to q_1$).

Indicele cu structură fixă al productivităţii muncii calculat prin metoda inversă va folosi, prin analogie, ca ponderi valorile factorului cantitativ din perioada curentă (q_1) calculîndu-se cu una din formulele :

$$I_{1/0}^{\overline{w}(t)} = \frac{\Sigma t_0 q_1}{\Sigma q_1} : \frac{\Sigma t_1 q_1}{\Sigma q_1} = \overset{(1)}{\frac{\Sigma t_0 q_1}{\Sigma t_1 q_1}} = \overset{(2)}{\frac{\Sigma i_{1/0}^{w} T_1}{\Sigma T_1}} \overset{(3)}{}. \qquad (11.109)$$

Indicele cu structură fixă, sub formă de raport de medii, se va calcula cînd interesează să se estimeze şi consumul mediu de timp de muncă necesar producţiei curente dacă muncitorii lucrau cu normele de consum din perioada de bază. Indicele cu structură fixă sub formă de agregat se foloseşte atunci cînd indicele productivităţii muncii se include în sistemul indicilor cheltuielilor totale de timp de muncă şi prin diferenţa dintre numărător şi numitor se stabileşte economia (risipa) de timp de muncă în funcţie de creşterea productivităţii muncii :

$$\Delta_{1/0}^{T(t)} = \Sigma t_0 q_1 - \Sigma t_1 q_1. \qquad (11.110)$$

Această relaţie arată cît este de mare diferenţa dintre timpul care ar fi fost necesar pentru producţia curentă dacă productivitatea rămînea la nivelul perioadei de bază şi timpul total consumat efectiv.

Indicele aritmetic ponderat al productivităţii muncii se aplică atunci cînd se cunosc indicii individuali şi cheltuiala totală de timp de muncă din perioada curentă ($q_1 t_1 = T_1$).

Indicele de variaţie a structurii, în acest caz, presupune că este variabilă numai producţia, iar nivelul de productivitate exprimat prin cheltuielile de timp de muncă pe unitate de produs a rămas acelaşi, adică :

$$I_{1/0}^{\overline{w}(q)} = \frac{\Sigma t_0 q_0}{\Sigma q_0} : \frac{\Sigma t_0 q_1}{\Sigma q_1}. \qquad (11.111)$$

Şi aceşti trei indici (11.108 ; 11.109 ; 11.111) se găsesc într-un sistem în care indicele cu structură variabilă (11.108) se descompune în produsul indicilor factoriali (11.109), (11.111).

11.5. PARTICULARITĂŢI ÎN CONSTRUIREA UNOR INDICI

Varietatea aspectelor care trebuie să fie studiată cu ajutorul metodei indicilor face dificilă examinarea particularităţilor ce apar în construirea acestora. Au fost aleşi indicii planului şi indicii teritoriali datorită importanţei lor pentru statistica socialistă.

Scanned with OKEN Scanner

11.5.1. Indicii sarcinii de plan şi ai îndeplinirii planului

Dezvoltarea planificată a economiei naţionale în socialism impune în mod necesar elaborarea şi perfecţionarea continuă a metodologiei de calcul şi interpretare a sarcinilor de plan şi a gradului de îndeplinire a lor.

În acest context, indicii planului sînt de două feluri : ai sarcinilor de plan şi ai îndeplinirii planului. Aceşti indici trebuie să fie astfel elaboraţi încît să poată fi analizaţi împreună şi comparabili cu datele statistice din perioada de bază a planului.

Indicii planului se pot calcula ca indici individuali şi indici de grup, păstrîndu-se, în general, analogia cu indicii de dinamică ai acelorași fenomene. Şi în acest caz problema cea mai dificilă este aceea a alegerii ponderilor, astfel încît să asigure relaţiile cantitative dintre indicii sarcinii de plan şi ai îndeplinirii planului, dar mai ales să permită stabilirea unor legături cu locul pe care-l ocupă indicatorul în cadrul planului naţional unic de dezvoltare, la nivel micro sau macroeconomic.

Pentru exemplificare se vor prezenta cîţiva indici a căror elaborare necesită rezolvarea mai multor probleme metodologice.

În cazul indicilor volumului fizic ai sarcinii de plan şi ai îndeplinirii planului se porneşte de la următoarele date : cantitatea de produse din perioada de bază (q_0) ; cantitatea de produse planificată pentru perioada curentă (q_{pl}) şi cantitatea de produse realizată în perioada curentă (q_1).

Pe baza acestor elemente se calculează *indicii individuali* ai volumului fizic :

— indicele sarcinii de plan ($i^q_{pl/0}$) :

$$i^q_{pl/0} = \frac{q_{pl}}{q_0} \cdot 100 ; \qquad (11.112)$$

— indicele îndeplinirii planului ($i^q_{1/pl}$) :

$$i^q_{1/pl} = \frac{q_1}{q_{pl}} \cdot 100. \qquad (11.113)$$

Dacă se compară cele două formule se poate observa că au un element comun (q_{pl}), ceea ce înseamnă că se poate stabili o relaţie de sistem din care să rezulte indicele dinamicii fenomenului de indexat calculat faţă de perioada de bază a planului. Această relaţie se poate enunţa astfel : produsul dintre indicele sarcinii de plan şi a îndeplinirii planului este egal cu indicele dinamicii aceluiaşi fenomen calculat faţă de baza de calcul. Pe exemplu volumului fizic :

$$i^q_{pl/0} \times i^q_{1/pl} = i^q_{1/0} , \qquad (11.114)$$

adică :

$$\frac{q_{pl}}{q_0} \times \frac{q_1}{q_{pl}} = \frac{q_1}{q_0} .$$

În mod corespunzător se stabileşte relaţia dintre modificările absolute. Între aceste elemente de calcul există relaţia :

$$q_1 - q_0 = (q_{pl} - q_0) + (q_1 - q_{pl}). \qquad (11.115)$$

Scanned with OKEN Scanner

Indicii de grup ai planului volumului fizic se calculează folosind ca ponderi prețurile din perioada de bază, astfel:

— indicele de grup al sarcinii de plan ($I^q_{pl/0}$):

$$I^q_{pl/0} = \frac{\Sigma q_{pl} p_0}{\Sigma q_0 p_0} \cdot 100; \tag{11.116}$$

— indicele de grup al îndeplinirii planului ($I^q_{1/pl}$):

$$I^q_{1/pl} = \frac{\Sigma q_1 p_0}{\Sigma q_{pl} p_0} \cdot 100. \tag{11.117}$$

Făcînd produsul celor doi indici se obține indicele dinamicii producției din perioada curentă față de perioada de bază, adică:

$$\frac{\Sigma q_{pl} p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_0}{\Sigma q_{pl} p_0} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}. \tag{11.118}$$

Pentru analiza statistică se mai poate calcula și un indice al îndeplinirii planului volumului fizic folosind ca ponderi prețurile planificate:

$$I^q_{1/pl} = \frac{\Sigma q_1 p_{pl}}{\Sigma q_{pl} p_{pl}} \cdot 100. \tag{11.119}$$

Probleme mai dificile apar însă cînd indicele de grup se calculează folosind ponderile din perioada curentă, cum este cazul indicelui prețurilor sau al costurilor de producție.

Pentru exemplificare se va folosi sistemul de indici ai costurilor de producție. Costurile de producție exprimă influența concomitentă a costurilor pe fiecare produs, cît și a cantității de produse fabricate.

Dacă costul pe produs se notează cu c, atunci indicii individuali ai planului vor fi:

— indicele sarcinii de plan ($i^c_{pl/0}$):

$$i^c_{pl/0} = \frac{c_{pl}}{c_0} \cdot 100; \tag{11.120}$$

— indicele îndeplinirii planului ($i^c_{1/pl}$):

$$i^c_{1/pl} = \frac{c_1}{c_{pl}} \cdot 100. \tag{11.121}$$

Făcînd produsul lor se obține indicele individual al dinamicii costului pe produs:

$$\frac{c_{pl}}{c_0} \cdot \frac{c_1}{c_{pl}} = \frac{c_1}{c_0}. \tag{11.122}$$

Scanned with OKEN Scanner

Indicele de grup al costurilor se calculează folosind ca ponderi cantitatea de produse. Costul de producţie, fiind un indicator care face parte din sistemul de preţuri şi tarife al mărfurilor şi serviciilor realizate în economia naţională, trebuie să folosească ca ponderi cantităţile din perioada înscrisă în numărătorul indicelui. Aceasta înseamnă că pentru indicele de grup al sarcinii de plan se va pondera cu cantităţile planificate, adică :

$$I^c_{\text{pl}/0} = \frac{\Sigma\, q_{\text{pl}}\, c_{\text{pl}}}{\Sigma\, q_{\text{pl}}\, c_0} \cdot 100. \qquad (11.123)$$

Aceasta înseamnă că se raportează costurile de producţie planificată la cît ar fi trebuit să coste acelaşi volum de producţie, dar la costurile perioadei de bază.

Folosind ca pondere producţia planificată apare necesitatea de a delimita producţia comparabilă, deoarece la numitor ea trebuie să fie estimată la costurile din perioada de bază. Rezultă, de aici, că prin *producţie comparabilă* se vor înţelege acele produse care s-au fabricat şi în perioada de bază, deci, există fişă de calculaţie a costului pe produs.

Pentru a măsura gradul de îndeplinire a planului costurilor de producţie se pot folosi ca ponderi fie producţia realizată, fie cea planificată în aceeaşi perioadă.

Dacă se foloseşte ca pondere producţia planificată se consideră ca nivel al planului costurile planificate, care sînt condiţionate de o anumită producţie fizică faţă de care se pot calcula economiile sau risipa de cheltuieli de producţie. În acest caz, formula va fi :

$$I^c_{\text{1/pl}} = \frac{\Sigma\, q_{\text{pl}}\, c_1}{\Sigma q_{\text{pl}} c_{\text{pl}}} . \qquad (11.124)$$

Acest indice permite să se calculeze şi indicele dinamicii costurilor producţiei planificate, înmulţind indicele sarcinii de plan cu indicele îndeplinirii planului :

$$\frac{\Sigma\, q_{\text{pl}}\, c_{\text{pl}}}{\Sigma\, q_{\text{pl}}\, c_0} \times \frac{\Sigma\, q_{\text{pl}}\, c_1}{\Sigma\, q_{\text{pl}}\, c_{\text{pl}}} = \frac{\Sigma\, q_{\text{pl}}\, c_1}{\Sigma\, q_{\text{pl}}\, c_0} . \qquad (11.125)$$

Acest sistem de indici nu este suficient pentru a analiza gradul de îndeplinire a costurilor de producţie, care trebuie să ia în discuţie costurile efectiv realizate ($c_1 q_1$) şi să le raporteze la mărimea lor dacă producţia efectiv realizată s-ar fi produs la costurile planificate. De aici rezultă necesitatea de a calcula şi un indice al îndeplinirii costurilor de producţie în structura efectivă a producţiei, adică :

$$I^c_{\text{1/pl}} = \frac{\Sigma\, q_1\, c_1}{\Sigma\, q_1\, c_{\text{pl}}} \cdot 100. \qquad (11.126)$$

Pe baza indicilor prezentaţi se pot calcula şi modificările absolute corespunzătoare, făcîndu-se diferenţa între numărătorul şi numitorul indicelui şi, după semnul diferenţei, se interpretează gradul de depăşire a planului în funcţie de natura sarcinii de plan stabilită în raport cu conţinutul economic al indicatorului a cărei variaţie se măsoară.

Scanned with OKEN Scanner

În concluzie se poate spune, ca și în cazul indicilor planului, că formulele generale se particularizează ținînd seama de specificul fenomenului de indexat și de datele de care se dispune.

11.5.2. Indicii teritoriali

Problema măsurării și comparării pe plan teritorial a diferiților indicatori statistici a constituit o temă larg dezbătută în literatura statistică din țara noastră și din alte țări socialiste.

Interesul acordat indicilor teritoriali, în statistica socialistă, este legat de importanța deosebită pe care statul socialist o acordă politicii de repartizare proporțională a forțelor de producție sub aspect teritorial, ca parte integrantă a procesului de industrializare socialistă a țării.

Necesitatea caracterizării statistice a distribuțiilor teritoriale, a explicării diferențelor de nivel realizate în dezvoltarea anumitor fenomene pune în fața statisticii o serie de sarcini care solicită adaptarea și perfecționarea metodei statistice în funcție de particularitățile repartiției teritoriale a caracteristicilor cercetate. Fără îndoială, în cadrul acestora funcția indicilor de a măsura variația unor fenomene asemănătoare existente în aceeași unitate de timp, în spații diferite poate răspunde cel mai bine problemelor de cunoaștere puse în fața statisticii teritoriale. Acesta este și unul din motivele pentru care măsurarea și compararea interjudețeană a fenomenelor este tratată ca fiind legată, în special, de elaborarea unui sistem de indici teritoriali.

În literatura de specialitate, în legătură cu elaborarea indicilor teritoriali ai comparării nivelurilor atinse de unele caracteristici în distribuția lor în spațiu, în aceeași unitate de timp, se poartă discuții în legătură cu următoarele aspecte metodologice :

a) alegerea bazei de comparare (raportare) în construirea indicilor teritoriali ;

b) alegerea formulei de calcul în funcție de conținutul indicatorilor care intră în componența indicilor teritoriali ;

c) alegerea metodei de calcul și problema ponderilor.

Alegerea bazei de raportare a indicilor teritoriali este problema esențială a metodologiei de elaborare a lor, deoarece, ca la orice mărime relativă, rezultatul comparării depinde tocmai de baza de referință luată în calcul. Practic este destul de dificil să se găsească un singur criteriu determinant, ca în cazul indicilor dinamicii sau ai îndeplinirii planului. Structura economică diferită, gradul de înzestrare tehnică diferit, se regăsesc într-o structură diferită a producției și a populației ocupate pe județe. Aceasta face ca nu întotdeauna să se poată dispune de anumite criterii obiective care să permită cu ușurință determinarea județului considerat ca bază de raportare, în special la compararea nivelurilor unor caracteristici calitative. Este situația specifică a unei economii de un dinamism rapid, cum este al țării noastre, în care locul județelor în structura social-economică a țării se modifică permanent. În consecință același județ poate să nu fie utilizat drept bază de raportare pentru o perioadă mai mare de timp și,

Scanned with OKEN Scanner

deci, indicii teritoriali să nu poată fi corelaţi într-o măsură eficientă cu cei de dinamică a fenomenului studiat. Deci acest aspect nu poate să nu fie luat în considerare în cazul comparării unor unităţi mai complexe — judeţele unei ţări sau chiar economiile diferitelor ţări — care, de regulă, prezintă diferenţe mari atît în structura economică, cît şi sub raport industrial şi tehnic.

Problema este relativ mai simplă atunci cînd se compară acelaşi indicator la două întreprinderi sau regiuni şi nu intervin astfel dificultăţi de ponderare diferită. În acest caz, testul de reversibilitate a indicilor teritoriali funcţionează fără nici o dificultate.

Dacă, de exemplu, se studiază o variabilă x pentru două unităţi teritoriale notate cu x_A şi x_B, atunci se pot obţine doi indici individuali, reversibili între ei :

$$i^x_{A/B} = \frac{x_A}{x_B}\ ; \quad i^x_{B/A} = \frac{x_B}{x_A}, \tag{11.127}$$

în care :

$$i^x_{A/B} \cdot i^x_{B/A} = 1, \tag{11.128}$$

$$\frac{x_A}{x_B} \cdot \frac{x_B}{x_A} = 1.$$

Evident că acest principiu al reversibilităţii, care trebuie să stea la baza sistemului de elaborare a indicilor teritoriali, nu poate fi totdeauna respectat, mai ales atunci cînd, în calculul lor, se folosesc indici agregaţi care trebuie să prezinte variaţia interregională a mai multor unităţi şi unde apar, inevitabil, şi un alt element, cu funcţie de ponderare. În acelaşi timp, la alcătuirea sistemului de indici teritoriali trebuie respectate şi evidenţiate în modul cel mai real legăturile obiective existente între variaţia caracteristicilor ce intră în sistem şi care determină alegerea formulei, cît şi a ponderilor.

Un alt aspect al problemei este legat de natura indicatorului comparat şi care determină alegerea formulei de calcul. În funcţie de natura indicatorului comparat pot apărea mai multe situaţii privind indicii teritoriali :

a) indicii teritoriali ai unui factor cantitativ însumabil (de exemplu, numărul de muncitori) şi care, de regulă, în raport cu o caracteristică derivată are rol de pondere. Formula de calcul, în acest caz, este :

$$I'_{A/B} = \frac{\Sigma f_A}{\Sigma f_B}, \tag{11.129}$$

în care A, B reprezintă unităţile teritoriale comparate ;

b) indicii teritoriali ai unui factor cantitativ neînsumabil direct, în care este necesară folosirea unui alt element ca pondere (exemplu, volumul producţiei obţinute). Formula va fi :

$$I'_{A/B} = \frac{\Sigma f_A x}{\Sigma f_B x}. \tag{11.130}$$

Aici pot interveni mai multe variante în funcţie de ponderea aleasă.

Scanned with OKEN Scanner

Cel mai frecvent se folosesc indicii agregaţi cu ponderile ambelor unităţi teritoriale :

$$I'_{A/B} = \frac{\Sigma f_A x_A}{\Sigma f_B x_A}; \qquad I'_{A/B} = \frac{\Sigma f_A x_B}{\Sigma f_B x_B}; \tag{11.131}$$

$$I'_{B/A} = \frac{\Sigma x_B f_A}{\Sigma x_A f_A}; \qquad I'_{B/A} = \frac{\Sigma x_B f_B}{\Sigma x_A f_B}. \tag{11.132}$$

Dacă ponderea o constituie preţul, atunci se pot compara producţiile a două ţări, exprimate într-o monedă de circulaţie internaţională.

Pentru a reduce influenţa ponderii asupra valorii indicilor calculaţi pe plan teritorial, se foloseşte din ce în ce mai des indicele lui Fisher adaptat la variaţia în spaţiu a fenomenelor :

$$I'_{A/B} = \sqrt{\frac{\Sigma f_A x_A}{\Sigma f_B x_A} \cdot \frac{\Sigma f_A x_B}{\Sigma f_B x_B}}; \tag{11.133}$$

c) indicii teritoriali ai unui factor calitativ calculaţi fie sub formă de agregat, fie sub formă de raport de medii :

$$I^{\bar{x}}_{A/B} = \frac{\Sigma x_A f}{\Sigma x_B f}; \qquad I^{\bar{x}}_{A/B} = \frac{\Sigma x_A f_A}{\Sigma f_A} : \frac{\Sigma x_B f_B}{\Sigma f_B}. \tag{11.134}$$

Dificultăţile pe care le ridică alegerea ponderilor au făcut ca, ţinînd seama de specificul comparaţiilor teritoriale, pentru calculul indicilor teritoriali cei mai mulţi statisticieni să propună metoda coeficienţilor standard.

Problema metodei de alcătuire a indicilor bazaţi pe determinarea unei colectivităţi standard a fost larg dezbătută în literatura statistică de specialitate. Această metodă, folosită în special la studiul fenomenelor demografice în calculul variaţiei între grupe a unor factori calitativi (de exemplu, la analiza indicelui de mortalitate dintre grupe) a fost adaptată la necesităţile de calcul ale indicilor teritoriali.

Cele două variante de calcul în care se aplică metoda coeficienţilor standard derivă din formula unui indice general cu structură variabilă, de forma :

$$I = \frac{\Sigma x_1 l_1}{\Sigma x_0 l_0}, \tag{11.135}$$

de unde :

a) indicele coeficienţilor standard (metoda directă) :

$$I = \frac{\Sigma x_1 l}{\Sigma x_0 l}; \tag{11.136}$$

b) indicele coeficienţilor standard (metoda indirectă) :

$$I = \frac{\Sigma x_1 l_1}{\Sigma x_1 l_1} : \frac{\Sigma x_0 l_0}{\Sigma x_0 l_0}, \tag{11.137}$$

Scanned with OKEN Scanner

în care :

x_0x_1 reprezintă nivelurile caracteristicii pe grupe ;
l_0l_1 — ponderea grupei respective a colectivității ;
x,l — nivelul caracteristicii și ponderea pentru colectivitatea standard.

Aceste formule au fost adaptate la calculul indicilor teritoriali astfel :

— metoda directă a colectivității standard :

$$I^x_{A/B} = \frac{\Sigma x_A f}{\Sigma x_B f}; \tag{11.138}$$

— metoda indirectă a colectivității standard :

$$I^x_{A/B} = \frac{\Sigma x_A f_A}{\Sigma x f_A} : \frac{\Sigma x_B f_B}{\Sigma x f_B}, \tag{11.139}$$

în care :

f reprezintă ponderile colectivității standard ;
x — variabila x din colectivitatea standard.

Desigur, în analiza distribuțiilor teritoriale a comparării nivelurilor aceloraşi caracteristici existente în aceeaşi unitate de timp în spații diferite se pot utiliza un număr foarte mare de combinații statistice în funcție de baza de comparație aleasă, de formule de calcul şi ponderile utilizate. De aceea, efectuarea analizei calitative prealabile a fenomenelor este şi aici cu deosebire necesară pentru ca, ținînd seama de natura specifică a conținutului indicatorului, cît şi de forma sa de manifestare teritorială concretă, să se utilizeze metoda cea mai adecvată pentru elaborarea indicilor teritoriali care măsoară variația nivelurilor spațiale ale caracteristicilor cercetate.

Scanned with OKEN Scanner

PROBLEME METODOLOGICE ALE INTERPRETĂRII DATELOR STATISTICE

Interpretarea rezultatelor cercetării statistice constituie faza în care se desăvîrşeşte procesul cunoaşterii statistice şi se formulează concluziile teoretice şi practice din întregul material statistic obţinut din observare şi prelucrare. Se poate spune, deci, că această fază a cunoaşterii statistice este cea mai importantă, în care cercetătorul trebuie să folosească cu competenţă cunoştinţe atît din domeniul statisticii, cît şi al celorlalte discipline cu care vine în contact atunci cînd realizează procesul de cercetare. El trebuie să cunoască din punct de vedere teoretic legile care acţionează în mod obiectiv în cadrul fenomenelor şi proceselor sociale şi economice pe care le caracterizează sub raport statistic, precum şi noţiunile şi categoriile corespunzătoare conţinutului social-economic al acestora. De aici rezultă că în această fază, mai mult decît în cele precedente, se realizează legătura reciprocă dintre statistică şi celelalte ştiinţe şi, în primul rînd, legătura dintre statistică, materialismul istoric şi economia politică marxistă.

12.1. BAZELE TEORETICE ŞI SARCINILE INTERPRETĂRII DATELOR STATISTICE

Statistica studiază fenomene sociale şi economice de masă, care apar ca fenomene de tip colectiv a căror proprietăţi nu pot fi cunoscute decît la nivelul întregului ansamblu prin interpretarea statistică a datelor dintr-un număr mare de cazuri individuale diferite, dar aparţinînd aceleiaşi structuri social-economice. Complexitatea acestor fenomene impune cu necesitate folosirea combinată a unor metode şi procedee variate de calcul a căror rezultate sînt sintetizate într-un sistem de indicatori statistici, conceput pentru diferite grade de prelucrare statistică. În statistica din ţara noastră, la construirea sistemului de indicatori se au în vedere următoarele criterii principale de clasificare : ramura economiei naţionale, ramura producţiei materiale, forma de subordonare administrativă, forma de proprietate şi împărţire administrativ-teritorială. În acest sens, sistemul de indicatori ai economiei naţionale se compune din mai multe subsisteme, între care se pot stabili atît relaţii privite orizontal — la acelaşi nivel de prelucrare — cît şi vertical, între elementele corespondente ale sistemului general în subsistemele de prelucrare a datelor statistice. De remarcat este faptul că sistemul general de indicatori ai economiei naţionale nu constituie o sumă a subsistemelor obţinute pe ramură, forme de proprietate, forme de subordonare sau repartizare teritorial-administrative, ci o agregare a acestora, în care, pentru fiecare indicator sintetic, se obţine o distribuţie de valori de înregistrare la nivele corespunzătoare de prelucrare a datelor. Numai în acest fel se pot interpreta proporţiile şi

Scanned with OKEN Scanner

corelaţiile existente în interiorul economiei naţionale în fiecare etapă dată, pe baza cărora se elaborează în continuare diferite variante de prognoză.

Interpretarea rezultatelor prezentate sub formă de sisteme de indicatori generalizatori şi totalizatori este partea cea mai dificilă şi de mare răspundere a întregului proces de cercetare statistică. De aceea, analiza întregului material statistic obţinut din observare şi prelucrare nu poate duce la concluzii valoroase decît dacă se bazează pe principiile teoretice ale tuturor disciplinelor ştiinţifice care studiază acelaşi obiect. Prin aceasta nu trebuie înţeles că s-ar minimaliza rolul statisticii în realizarea procesului de cunoaştere al societăţii, ci doar faptul că i se recunosc limitele şi sînt precizate raporturile de interdependenţă cu celelalte ştiinţe.

Trebuie precizat, de asemenea, că procesul de interpretare ştiinţifică a datelor este de neconceput fără utilizarea largă a principiilor teoretice ale filozofiei marxiste şi a calculului probabilităţilor.

Fundamentarea cercetării statistice pe principiile dialecticii materialiste şi ale teoriei probabilităţilor a permis statisticii să treacă pragul de la descrierea cantitativă şi formularea unor regularităţi în producerea fenomenelor sociale, la analiza lor, la cunoaşterea legilor care le determină şi, pe această bază, la efectuarea de previziuni ştiinţifice.

Interpretarea datelor statistice este deci o muncă complexă, în care, odată cu munca propriu-zisă de statistică se realizează şi folosirea cunoştinţelor puse la îndemînă de alte ştiinţe şi care condiţionează toate etapele cercetării statistice.

Interpretarea rezultatelor cercetării statistice nu poate fi ruptă de întreaga muncă de analiză, ce se efectuează pe tot parcursul cunoaşterii statistice. Mai mult, se poate spune că această ultimă fază a cercetării statistice este condiţionată de modul în care s-au efectuat pregătirea şi desfăşurarea observării şi prelucrării statistice. Este suficient să se amintească aici, că orice cercetare statistică începe printr-o muncă de analiză teoretică, pe baza căreia se întocmeşte planul cercetării, se definesc obiectul şi unitatea de observare, se stabileşte programul de observare şi prelucrare. Analiza este cu atît mai necesară în procesul de prelucrare, pentru alegerea metodelor şi procedeelor de prelucrare statistică în raport cu conţinutul specific al fenomenelor şi al relaţiilor obiective de cauzalitate dintre caracteristicile studiate. Interpretarea rezultatelor în vederea formulării concluziei statistice nu apare decît ca o sinteză a analizelor parţiale, efectuate pe tot parcursul cunoaşterii statistice, în care se confruntă şi se compară rezultatele parţiale în scopul generalizării lor.

Rezultă că, la formularea concluziilor statistice, trebuie să se folosească combinat rezultatele obţinute în toate fazele cercetării, să se verifice permanent autenticitatea datelor obţinute şi limitele în care ele pot fi utilizate. Aceasta înseamnă efectuarea unei analize atente a materialului statistic pe tot parcursul cercetării. Sintetizînd şi interpretînd toate aceste elemente, în etapa finală a procesului de cercetare se efectuează întreaga asamblare a rezultatelor care are ca scop caracterizarea sub raport statistic a colectivităţii studiate. Acesta este unul din motivele pentru care această etapă se mai numeşte şi analiza statistică.

Se poate spune, deci, că problemele analizei statistice au fost puse în cea mai mare parte în capitolele precedente ale manualului

Scanned with OKEN Scanner

şi că ceea ce rămîne de făcut este de a le cuprinde acum, pe baza cunoştinţelor obţinute, într-o prezentare sistematică, abordînd şi unele aspecte noi, specifice acestei etape finale.

Pe lîngă aceasta trebuie să se atragă atenţia asupra unui fapt cunoscut, anume că dacă este valabilă afirmaţia că metodologia statistică nu poate da reţete pentru rezolvarea nenumăratelor situaţii concrete pe care le oferă cercetarea statistică, atunci acest lucru este cu atît mai valabil pentru faza de conturare a concluziilor statistice. Rolul statisticii nu este de a prezenta astfel de reţete, ci de a se călăuzi după principiile pe care trebuie să le respecte orice interpretare ştiinţifică a unui bogat material faptic şi de a alege şi utiliza cele mai potrivite metode şi procedee în analiza concretă a fenomenelor şi proceselor social-economice.

Prima problemă care se ridică pentru interpretarea datelor statistice este aceea a formulării sarcinilor de cunoaştere pe care trebuie să le rezolve. Sarcinile analizei statistice se stabilesc pentru fiecare cercetare statistică în parte; formularea cea mai largă a acestora s-a făcut atunci cînd s-au stabilit scopul şi obiectul observării. Aceste sarcini rămîn valabile şi pentru munca de interpretare a rezultatelor, cu modificările pe care le poate aduce materialul concret obţinut în urma observării şi prelucrării.

Într-o formă generală principalele sarcini ale interpretării datelor statistice trebuie să aibă în vedere următoarele aspecte:

— analiza fenomenelor supuse cercetării, pe baza sistemului de indicatori, totalizatori şi sintetici, ce se obţin la nivelul întregului ansamblu;

— analiza structurii şi a deplasărilor de structură intervenite de la o perioadă de timp la alta, pe tipurile calitative existente în cadrul colectivităţii cercetate;

— analiza valorilor tipice pe fiecare grupă şi pe total şi a legăturilor dintre ele pe baza indicatorilor medii şi a indicatorilor de variaţie şi asimetrie;

— analiza legăturilor de interdependenţă dintre fenomene şi dintre caracteristicile statistice folosind indicatorii simpli şi sintetici de corelaţie şi regresie;

— analiza în dinamică a fenomenelor în vederea desprinderii tendinţei obiective de dezvoltare şi a reliefării influenţei diferiţilor factori asupra fenomenelor complexe în scopul elaborării diferitelor variante de prognoză;

— verificarea ipotezelor, estimarea parametrilor colectivităţii generale pe baza analizei datelor de selecţie;

— formularea concluziilor statistice; expunerea analizei.

Aceste probleme pot să apară toate într-o singură cercetare statistică de mare amploare sau pot să apară numai unele în funcţie de scopul cercetării şi de datele disponibile.

De aceea, în cercetarea concretă, precizarea sarcinilor interpretării materialului statistic se face de fiecare dată în raport cu scopul urmărit. Analiza statistică poate avea drept scop o caracterizare multiplă a fenomenului sau a procesului social-economic, sau numai caracterizarea unuia din laturile lui, de a desprinde anumite aspecte. Şi într-un caz şi în altul trebuie formulate clar sarcinile analizei. Precizarea incompletă a sarcini-

Scanned with OKEN Scanner

lor analizei, formularea lor confuză poate duce la alunecarea întregii analize spre alte obiective decît cele urmărite, împiedicînd obținerea rezultatelor dorite în funcție de sarcinile care i-au fost puse.

12.2. SURSELE DE DATE UTILIZATE ÎN ANALIZA STATISTICĂ

În cercetarea statistică, după cum se știe, se pot folosi date pe care statistica și le obține în mod nemijlocit prin organizarea unor observări statistice speciale sau poate să folosească date obținute dintr-o înregistrare anterioară pe baza diferitelor documente de evidență. De asemenea, se pot face analize de date statistice din cercetări anterioare prezentate în diferite publicații. Fără îndoială, că interpretarea datelor statistice trebuie să țină seama de sursa de informație statistică pe care a utilizat-o. În cazul în care datele provin dintr-o cercetare statistică special organizată, planul de cercetare include de la început și principalele probleme ale interpretării materialului statistic, care va fi obținut în urma efectuării observării și a prelucrării statistice. Probleme deosebite apar, în special, atunci cînd se utilizează date de la mai multe cercetări statistice sau din surse diferite. În acest caz, problema esențială este aceea a asigurării posibilităților de comparare a acestor date.

Dificultăți pentru munca de interpretare apar și atunci cînd datele utilizate sînt obținute din observări statistice parțiale. În acest caz trebuie să se verifice, în primul rînd, care este gradul de reprezentativitate a datelor parțiale și numai în măsura în care partea supusă observării îndeplinește într-un grad suficient de mare această condiție, prin interpretarea științifică a datelor parțiale, se vor putea stabili limitele intervalului de încredere pentru caracterizarea întregului ansamblu.

În esență, examinarea atentă a surselor de date folosite în analiza statistică, înseamnă o măsură necesară pentru asigurarea comparabilității datelor și pentru aprecierea limitelor lor în caracterizarea fenomenelor studiate.

12.3. METODE ȘI PROCEDEE DE INTERPRETARE A DATELOR STATISTICE

Printre metodele și procedeele de interpretare a datelor statistice se pot aminti :

— verificarea autenticității datelor ;
— confruntarea și compararea datelor ;
— gruparea pe tipuri calitative și calcularea indicatorilor derivați ce caracterizează fiecare grupă și care permit stabilirea locului pe care îl ocupă în întreaga colectivitate ;
— metode de analiză a distribuțiilor de frecvență ;
— metoda de analiză a seriilor statistice interdependente ;
— metode statistice de analiză a seriilor cronologice ;
— metode de analiză statistică a repartiției teritoriale ;

Scanned with OKEN Scanner

— metode de extindere a datelor de selecţie asupra întregului ansamblu.

Verificarea autenticităţii datelor presupune efectuarea unui nou control asupra modului în care au fost obţinute datele, precum şi asupra modului de prelucrare a acestora.

Verificarea datelor statistice trebuie să aibă în vedere atît aspectele cantitative, dar mai ales verificarea modului în care s-a asigurat un conţinut real materialului faptic de care se dispune. Aceasta se poate realiza prin verificarea modului în care elementele planului de observare — obiect, unitatea de observare, programul observării, timpul observării etc. — au fost respectate, precum şi măsura în care prelucrarea statistică s-a făcut în mod diferenţiat prin folosirea celor mai adecvate metode şi procedee de calcul statistic.

O atenţie deosebită trebuie acordată verificării grupărilor utilizate în prelucrarea, respectiv dacă alegerea caracteristicilor s-a făcut în mod corespunzător, dacă numărul de grupe a permis studierea formei de variaţie a caracteristicii, dacă intervalele de grupare au fost bine stabilite.

Se poate spune că acum se face o adevărată verificare a calităţii datelor statistice, a corespondenţei lor cu realitatea.

O atenţie deosebită trebuie acordată examinării critice a materialelor obţinute prin cercetările monografice şi selective. Examinarea critică a acestora trebuie să arate dacă ele se pot extinde asupra colectivităţii generale şi sînt suficiente pentru reliefarea tipurilor social-economice ale fenomenelor cercetate. Trebuie să se verifice în toate cazurile reprezentativitatea datelor obţinute prin observarea selectivă, să se întrebuinţeze toate mijloacele pentru stabilirea valabilităţii extinderii concluziilor asupra colectivităţii generale.

Confruntarea şi compararea datelor reprezintă procedee de bază în munca de interpretare statistică a rezultatelor obţinute din observare şi prelucrare. Datele statistice care se pot obţine prin mai multe căi sînt confruntate pentru a li se verifica exactitatea. De exemplu, dările de seamă statistice ajung la Direcţia Centrală de Statistică prin două căi : prin centralizarea indicatorilor de către organele sale teritoriale (direcţiile judeţene de statistică) şi prin darea de seamă centralizată de fiecare minister sau organizaţie centrală. Confruntînd cele două dări de seamă centralizate se poate verifica existenţa eventualelor erori care vor fi înlăturate odată cu depistarea locului şi cauzei.

Pe lîngă confruntare, în analiza statistică se utilizează pe larg compararea datelor statistice fie pentru analiza desfăşurării fenomenelor în timp, fie pentru compararea diferitelor fenomene sau laturi ale fenomenelor coexistente în spaţiu.

O primă problemă care trebuie rezolvată este aceea a verificării posibilităţii de comparare directă a datelor. Aceasta presupune o analiză a conţinutului indicatorilor şi formei de exprimare a acestora, a metodelor de observare şi calcul statistic care au fost utilizate.

Studiind principalele cauze care pot face ca unele date statistice să nu fie comparabile, se ajunge la concluzia că ele pot fi reduse la următoarele :

— schimbări teritoriale ;
— schimbări organizatorice ;

— definirea neclară sau nejustă a obiectului observării statistice;
— procedee diferite de observare;
— culegerea de date din diferite surse;
— perioade sau momente diferite ale observării;
— unități de măsură sau coeficienți de echivalență diferiți;
— schimbarea prețurilor de la o perioadă la alta.

Aceste probleme au fost discutate în legătură cu diferitele metode de calcul ale indicatorilor derivați, prezentate în cadrul cursului de statistică și, cu acest prilej, s-a arătat și modul în care ele se pot înlătura în vederea asigurării comparabilității datelor. S-a precizat, de asemenea, că dacă nu există posibilități de a face datele comparabile, nu se poate trece la calcularea indicatorilor derivați, care, atunci cînd provin din date necomparabile între ele, nu au un conținut real și reflectă denaturat fenomenele cercetate.

În concluzie, în cazul în care una din aceste cauze nu permite compararea directă a datelor, trebuie să se analizeze dacă este posibilă o reconstituire a seriilor de date statistice, care să permită interpretarea corectă a rezultatelor ce se vor obține pe baza noilor serii.

Gruparea pe tipuri calitative se poate realiza în urma interpretării seriilor statistice și a indicatorilor derivați care au fost obținuți din prelucrarea lor pentru un număr suficient de caracteristici ce se găsesc într-o relație de interdependență. Verificarea omogenității grupelor, după cum s-a arătat, se realizează în statistică prin calculul indicatorilor medii și al celor de variație. Pe baza interpretării acestora se pot separa colectivitățile statistice pe grupe omogene, corespunzătoare tipurilor social-economice ce s-au conturat în mod obiectiv în colectivitatea supusă cercetării și care, într-un anumit sens, prezintă aceeași tendință de covariație. De aceea, de cele mai multe ori, gruparea pe tipuri calitative nu se poate realiza decît ca o grupare repetată (sau regrupare) în care se trece de la interpretarea independentă a fenomenelor la caracterizarea interdependentă dintre ele.

Gruparea repetată se utilizează, dealtfel, pentru a descoperi legătura dintre mai multe caracteristici statistice, prezentate sub forma unor serii paralele interdependente. Este indicat, de aceea, ca la folosirea unor asemenea serii să se facă distincție între caracteristicile factoriale și cele rezultative, astfel încît ele să corespundă, de fapt, unor grupări combinate, în care analiza raporturilor de corelație să se poată realiza cu ușurință.

Combinînd toate aceste rezultate obținute pe fiecare grupă tipică și calculînd, în continuare, indicatorii de structură se poate preciza care este locul pe care fiecare grupă îl ocupă în cadrul colectivității; comparațiile în dinamică sau pe plan teritorial pot aprecia mutațiile de structură intervenite în fenomenele studiate.

Metodele de analiză a distribuțiilor de frecvențe au fost prezentate pe larg într-un capitol al cursului Teoria Statisticii. Aici se va aminti numai importanța acestor valori tipice pentru cunoașterea legităților statistice și a separării influenței factorilor esențiali de cei neesențiali.

După cum s-a mai arătat, legitatea statistică se manifestă numai în cadrul unui număr mare de cazuri individuale diferite sub forma unei tendințe, în care abaterile într-un sens sau altul se compensează reciproc. Interpretarea acestor mărimi trebuie să se facă însă în raport cu gradul de omogenitate al colectivității sau al grupei pentru care aceste valori au

Scanned with OKEN Scanner

fost calculate şi, deci, interpretarea statistică trebuie să aibă în vedere modul în care se distribuie aceste abateri în raport cu probabilitatea lor de apariţie. În procesul de interpretare statistică a seriilor de distribuţie se realizează legătura dintre cunoştinţele teoretice şi modul concret de manifestare a acestor legităţi, care nu poate avea loc decît în această etapă finală cînd se cunosc toate rezultatele observării şi prelucrării.

Metodele de analiză a seriilor statistice interdependente reprezintă conţinutul principal al muncii de interpretare ştiinţifică a datelor statistice. Fundamentarea statisticii pe principiile dialecticii materialiste impune, în mod necesar, aplicarea legăturilor de cauzalitate dintre fenomenele şi procesele sociale şi economice ce intră în domeniul de cercetare al statisticii.

Fenomenele vieţii sociale şi economice nu pot fi cunoscute şi înţelese în afara legăturilor reciproce şi a condiţiilor care le determină dezvoltarea. De aceea, analiza care nu aplică cu pricepere trăsăturile fundamentale ale dialecticii materialiste nu poate să-şi îndeplinească sarcinile şi rămîne un simplu exerciţiu formal. În acelaşi timp trebuie arătat că nu pot fi exprimate întotdeauna şi în întregime, în datele statistice, caracterul legăturii reciproce şi legităţile de dezvoltare a tuturor fenomenelor şi proceselor sociale. Astfel, de exemplu, nu toate fenomenele se pot cuantifica şi, deci, supune unui calcul statistic. De aceea, aici se pune problema deosebit de importantă pentru cunoaşterea statistică, anume de a stabili fenomenele şi procesele sociale şi economice ale căror legături reciproce le pot exprima şi măsura, situaţiile în care poate face acest lucru, precum şi posibilităţile în care metodele statistice se pot extinde şi la fenomene care, prin natura lor, nu pot fi cercetate direct prin astfel de metode. Problema a fost pusă şi lămurită în capitolul despre obiectul statisticii şi în capitolul privind analiza legăturilor de corelaţie dintre fenomene. Tot cu acest prilej a fost lămurită şi problema posibilităţilor şi condiţiilor în care se pot utiliza instrumentele statistico-matematice în cercetarea concretă a fenomenelor.

Metode statistice de analiză a seriilor cronologice. Una din sarcinile principale ale statisticii o constituie prezentarea datelor statistice sub formă de serii cronologice (dinamice).

În legătură cu stabilirea unor astfel de serii statistice, alegerea bazei de comparaţie şi a etapei pentru care se întocmeşte seria prezintă o importanţă excepţională la justa caracterizare a fenomenelor social-economice, pentru determinarea particularităţilor dezvoltării lor. În funcţie de mărimea seriei se pot aprecia tendinţele de lungă durată ce se manifestă în cadrul fenomenelor cercetate şi modificarea acestora de la o etapă la alta.

Interpretarea corectă a tendinţei obiective după care s-au dezvoltat fenomenele studiate, permite alegerea formulelor de calcul pentru analiza dezvoltării lor în perspectivă. Astfel, o analiză amănunţită şi corectă a unei serii statistice cronologice poate servi ca bază reală în elaborarea sarcinilor de plan în etapa care urmează. De asemenea, pe baza interpretării indicatorilor obţinuţi din prelucrarea unei serii cronologice se poate stabili care a fost raportul dintre acţiunea factorilor esenţiali şi celor întîmplători în determinarea fenomenului studiat, modul de acţiune a factorilor esen-

Scanned with OKEN Scanner

ţiali exprimat sub forma unei tendinţe rectilinii sau curbilinii, păstrarea acestei tendinţe pe toată etapa sau modificarea ei chiar în interiorul acesteia etc.

Folosirea în analiză a unor serii dinamice paralele dă posibilitatea de a compara nu numai modificările unei singure caracteristici, ci de a sesiza şi exprima numeric direcţia, tendinţa acestei schimbări pentru mai multe caracteristici, precum şi corelaţiile şi proporţiile ce se pot realiza în dinamică.

Metodele de analiză statistică a repartiţiilor teritoriale se folosesc pentru a analiza variaţia fenomenelor sociale şi economice în fiecare etapă pe forme de organizare administrativ-teritoriale. Pentru aceasta este necesar să se identifice şi ierarhizeze caracteristicile economice, sociale, demografice şi culturale care definesc locul pe care îl ocupă fiecare judeţ în economia ţării. De asemenea, este necesar să se analizeze care este gradul de dispersie a acestor caracteristici; în general, caracteristicile care reflectă nivelul de trai al populaţiei în economia socialistă trebuie să prezinte o dispersie foarte mică, în timp ce celelalte caracteristici, reflectînd particularităţile teritoriale, vor avea o dispersie mai mare.

În analiza statistică pe plan teritorial trebuie să se folosească combinat toate metodele statistice, prin care să se caracterizeze atît variaţia independentă a fenomenelor, cît şi interdependenţa dintre ele.

Metode de analiză factorială. În formarea fenomenelor complexe, studiate de statistică, nu toţi factorii care le determină au acelaşi rol şi acelaşi grad de intensitate. De asemenea, aceşti factori pot acţiona în acelaşi sens sau în direcţii diferite. Pentru a putea stabili care este gradul de influenţă a diferiţilor factori este necesar să se descompună mărimea fenomenului complex în funcţie de mai multe variabile, corespunzătoare caracteristicilor-factori. În interpretarea gradului de influenţă a caracteristicilor-factori cel mai frecvent se aplică metoda corelaţiei şî metoda indicilor. Interpretarea statistică a rezultatelor permite să se evidenţieze gradul de influenţă a acestor factori şi să se estimeze posibilităţile de dezvoltare a fenomenului complex cercetat în etapa care urmează.

De aceea, analiza factorială bazată pe interpretarea datelor statistice îşi găseşte o largă aplicabilitate în studiul modului de îndeplinire a sarcinilor de plan atît în cadrul fiecărei întreprinderi, cît şi la nivelul întregii ramuri. Rezultatele aplicării analizei factoriale sînt folosite pentru fundamentarea sarcinilor de plan privind dezvoltarea fenomenelor complexe.

Metodele de extindere a datelor de selecţie asupra întregului ansamblu îşi găsesc din ce în ce mai frecvent aplicarea în cercetarea statistică concretă. De regulă, cercetarea selectivă se realizează pentru un număr mare de caracteristici statistice pe baza cărora se poate calcula întreaga gamă de indicatori obţinuţi prin aplicarea calculelor statistice. De aceea, la analiza şi folosirea datelor de selecţie se pun probleme legate de : verificarea gradului de reprezentativitate; calcularea mărimii erorii de reprezentativitate pentru caracteristicile la care se cunosc şi datele din observarea totală, precum şi estimarea erorilor probabile pentru caracteristicile la care nu se cunosc decît date de selecţie. La extinderea datelor selecţiei se pune problema de a estima parametrii colectivităţii totale calculaţi ca medii, dispersii, indici, coeficienţi de corelaţie etc. De asemenea, pe baza datelor de selecţie se verifică ipotezele cu privire la gradul de reprezentativitate

Scanned with OKEN Scanner

al mediei în raport cu distribuţia de selecţie, gradul de stabilitate al mediei şi dispersiei de la o selecţie la alta, gradul de semnificaţie al diferiţilor indicatori de corelaţie. În această etapă finală se folosesc toate rezultatele parţiale obţinute din aplicarea metodei selective în vederea conturării imaginii colectivităţii studiate.

12.4. EXPUNEREA REZULTATELOR INTERPRETĂRII DATELOR STATISTICE

Interpretarea datelor statistice, ca fază ultimă a cunoaşterii statistice, este aceea în care se formulează concluziile şi propunerile practice, care trebuie să cuprindă răspunsuri clare, complete şi exacte la obiectivele care au fost puse în faţa cercetării. Aceasta înseamnă că formularea concluziilor trebuie să fie bazată numai pe date statistice autentice şi ireproşabile, care nu provoacă nici o îndoială. Ele trebuie să fie confirmate de calculele economico-statistice de verificare; indicatorii derivaţi trebuie să se bazeze pe materiale statistice autentice, obţinute în mod ştiinţific şi să reprezinte temeiul pentru o serie de ipoteze statistice cu privire la modul de dezvoltare a fenomenelor în etapa viitoare.

Rezultatele interpretării datelor statistice se concretizează în anumite modalităţi de expunere, care îmbracă forma unei note analitice sau forma unei lucrări ştiinţifice.

Ori care ar fi forma de prezentare a rezultatelor analizei este necesar să se folosească ca material ilustrativ şi de fundamentare tabele şi grafice. Ele pot fi incluse în text sau sub formă de anexe la raportul de analiză.

Expunerea prin texte nu trebuie să fie însă prea întinsă, ci trebuie să se concentreze în jurul scopului principal scoţînd în evidenţă ceea ce este esenţial pentru fenomenele sau procesele cercetate. Pentru aceasta este necesar ca expunerea rezultatelor să se facă după un plan bine determinat, care este parte integrantă din planul cercetării.

Necesitatea întocmirii prealabile a întregului plan al cercetării statistice, unitatea sa organică în toate fazele asigură, pe lîngă autenticitatea cunoaşterii statistice, şi utilizarea raţională a acestui mijloc de cunoaştere care implică mari cheltuieli de forţă de muncă şi materiale.

Scanned with OKEN Scanner

Numere întîmplătoare

Seria	1	2	3	4	5	6	7	8
1	78 994	36 244	02 673	25 475	84 953	61 793	50 243	63 423
2	04 909	58 485	70 786	93 930	34 880	73 059	06 823	80 257
3	46 582	73 570	33 004	51 795	86 477	46 736	60 460	70 345
4	29 242	89 792	88 634	60 285	07 190	07 795	27 011	85 941
5	68 104	81 339	97 090	20 601	78 940	20 228	22 803	96 070
6	17 156	02 182	82 504	19 880	93 747	80 910	78 260	25 136
7	50 711	94 789	07 171	02 103	99 057	98 775	37 997	18 325
8	39 449	52 409	75 095	77 720	39 729	03 205	09 313	43 545
9	75 629	82 729	76 916	72 657	58 992	32 576	01 154	84 890
10	01 020	55 151	36 132	51 971	32 155	60 735	64 867	35 424
11	08 337	89 989	24 290	08 618	66 798	25 889	52 860	57 375
12	76 829	47 229	19 706	30 094	69 430	92 399	98 749	22 081
13	39 708	30 641	21 267	56 501	95 182	72 442	21 445	17 276
14	89 836	55 817	56 747	75 195	06 818	83 043	47 403	58 266
15	25 903	61 370	66 081	54 076	67 442	52 964	23 823	02 718
16	71 345	03 422	01 015	68 025	19 703	77 313	04 555	83 425
17	61 454	92 263	14 647	08 473	34 124	10 740	40 839	05 620
18	80 376	08 909	30 470	40 200	46 558	61 742	11 643	92 121
19	45 144	54 373	05 505	90 074	24 783	86 299	20 900	15 144
20	12 191	88 527	58 852	51 175	11 534	87 218	04 876	85 584
21	62 936	59 120	73 957	35 969	21 598	47 287	39 394	08 778
22	31 588	96 798	43 668	12 611	01 714	77 266	55 079	24 690
23	20 787	96 048	84 726	17 512	39 450	43 618	30 629	24 356
24	45 603	00 745	84 635	43 079	52 724	14 262	05 750	89 373
25	31 606	64 782	34 027	56 734	09 365	20 008	93 559	78 384
26	10 452	33 074	76 718	99 556	16 026	00 013	78 411	95 107
27	37 016	64 633	67 301	50 949	91 298	74 968	73 631	57 397
28	66 725	97 865	25 409	37 498	00 816	99 262	14 471	10 232
29	07 380	74 438	82 120	17 890	40 963	55 757	13 492	68 294
30	71 621	57 688	58 256	47 702	74 724	89 419	08 025	68 519
31	03 466	13 263	23 917	20 417	11 315	52 805	33 072	07 723
32	12 692	32 931	97 387	34 822	53 775	91 674	76 549	37 635
33	52 192	30 941	44 998	17 833	94 563	23 062	95 725	38 463
34	56 691	72 529	66 063	73 570	86 860	68 125	40 436	31 303
35	74 952	43 041	58 869	15 677	78 598	43 520	97 521	83 248
36	18 752	43 693	32 867	53 017	22 661	39 610	03 796	02 622
37	61 691	04 944	43 111	28 325	82 319	65 589	66 048	98 498
38	49 197	63 948	38 947	60 207	70 667	39 843	60 607	15 328
39	19 436	87 291	71 684	74 859	76 501	93 456	95 714	92 518
40	39 143	64 893	14 606	13 543	09 621	68 301	69 817	52 140
41	82 244	67 549	76 491	09 761	74 494	91 307	64 222	66 592
42	55 847	56 155	42 878	23 708	97 999	40 131	52 360	90 398
43	94 095	95 970	07 826	25 991	37 584	56 966	68 623	83 454
44	11 751	69 469	25 521	44 097	07 511	88 976	30 122	67 542
45	69 902	08 995	27 821	11 758	64 989	61 902	32 121	28 165
46	21 850	25 352	25 556	92 161	23 592	43 294	10 479	37 879
47	75 850	46 992	25 165	55 906	62 339	88 958	91 717	15 756
48	29 648	22 086	42 581	85 677	20 251	39 641	65 786	80 689
49	82 740	28 443	42 734	25 518	82 827	35 825	90 288	32 911
50	36 842	42 092	52 075	83 926	42 875	71 500	69 216	01 350

Reprodus după F. Mills, *Statistical Methode*, Columbia University, Third edition, New York, 1955.

Scanned with OKEN Scanner

Tabel cu valorile funcției Gauss-Laplace

$$\Phi_{(t)} = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt$$

t	$\Phi_{(t)}$	t	$\Phi_{(t)}$	t	$\Phi_{(t)}$	t	$\Phi_{(t)}$	t	$\Phi_{(t)}$
0,00	0,0000	0,30	0,2358	0,60	0,4515	0,90	0,6319	1,20	0,7699
0,01	0,0080	0,31	0,2434	0,61	0,4581	0,91	0,6372	1,21	0,7737
0,02	0,0160	0,32	0,2510	0,62	0,4647	0,92	0,6424	1,22	0,7775
0,03	0,0239	0,33	0,2586	0,63	0,4713	0,93	0,6476	1,23	0,7813
0,04	0,0319	0,34	0,2661	0,64	0,4778	0,94	0,6528	1,24	0,7850
0,05	0,0399	0,35	0,2737	0,65	0,4843	0,95	0,6579	1,25	0,7887
0,06	0,0478	0,36	0,2812	0,66	0,4907	0,96	0,6629	1,26	0,7923
0,07	0,0558	0,37	0,2886	0,67	0,4971	0,97	0,6680	1,27	0,7959
0,08	0,0638	0,38	0,2961	0,68	0,5035	0,98	0,6729	1,28	0,7995
0,09	0,0717	0,39	0,3035	0,69	0,5098	0,99	0,6778	1,29	0,8030
0,10	0,0797	0,40	0,3108	0,70	0,5161	1,00	0,6827	1,30	0,8064
0,11	0,0876	0,41	0,3182	0,71	0,5223	1,01	0,6875	1,31	0,8098
0,12	0,0955	0,42	0,3255	0,72	0,5285	1,02	0,6923	1,32	0,8132
0,13	0,1034	0,43	0,3328	0,73	0,5346	1,03	0,6970	1,33	0,8165
0,14	0,1113	0,44	0,3401	0,74	0,5407	1,04	0,7017	1,34	0,8198
0,15	0,1192	0,45	0,3473	0,75	0,5467	1,05	0,7063	1,35	0,8230
0,16	0,1271	0,46	0,3545	0,76	0,5527	1,06	0,7109	1,36	0,8262
0,17	0,1350	0,47	0,3626	0,77	0,5587	1,07	0,7154	1,37	0,8293
0,18	0,1428	0,48	0,3688	0,78	0,5646	1,08	0,7199	1,38	0,8324
0,19	0,1507	0,49	0,3759	0,79	0,5705	1,09	0,7243	1,39	0,8355
0,20	0,1585	0,50	0,3829	0,80	0,5763	1,10	0,7287	1,40	0,8385
0,21	0,1663	0,51	0,3899	0,81	0,5821	1,11	0,7330	1,41	0,8415
0,22	0,1741	0,52	0,3969	0,82	0,5878	1,12	0,7375	1,42	0,8444
0,23	0,1819	0,53	0,4039	0,83	0,5935	1,13	0,7415	1,43	0,8473
0,24	0,1897	0,54	0,4108	0,84	0,5991	1,14	0,7457	1,44	0,8501
0,25	0,1974	0,55	0,4177	0,85	0,6047	1,15	0,7499	1,45	0,8529
0,26	0,2051	0,56	0,4245	0,86	0,6102	1,16	0,7540	1,46	0,8557
0,27	0,2128	0,57	0,4313	0,87	0,6157	1,17	0,7560	1,47	0,8584
0,28	0,2205	0,58	0,4331	0,88	0,6211	1,18	0,7620	1,48	0,8611
0,29	0,2282	0,59	0,4448	0,89	0,6265	1,19	0,7660	1,49	0,8638
									(continuare)
1,50	0,8664	1,72	0,9146	1,94	0,9476	2,32	0,9797	2,76	0,9942
1,51	0,8690	1,73	0,9164	1,95	0,9488	2,34	0,9807	2,78	0,9946
1,52	0,8715	1,74	0,9181	1,96	0,9500	2,36	0,9817	2,80	0,9949
1,53	0,8740	1,75	0,9199	1,97	0,9512	2,38	0,9827	2,82	0,9952
1,54	0,8764	1,76	0,9216	1,98	0,9523	2,40	0,9836	2,84	0,9955
1,55	0,8789	1,77	0,9233	1,99	0,9534	2,42	0,9845	2,86	0,9958
1,56	0,8812	1,78	0,9249	2,00	0,9545	2,44	0,9853	2,88	0,9960
1,57	0,8836	1,79	0,9265	2,02	0,9566	2,46	0,9861	2,90	0,9962
1,58	0,8859	1,80	0,9281	2,04	0,9587	2,48	0,9869	2,92	0,9965
1,59	0,8882	1,81	0,9297	2,06	0,9606	2,50	0,9876	2,94	0,9967
1,60	0,8904	1,82	0,9312	2,08	0,9625	2,52	0,9883	2,90	0,9969
1,61	0,8926	1,83	0,9328	2,10	0,9643	2,54	0,9889	2,98	0,9971
1,62	0,8948	1,84	0,9342	2,12	0,9660	2,56	0,9895	3,00	0,9973
1,63	0,8969	1,85	0,9357	2,14	0,9676	2,58	0,9901	3,20	0,9986
1,64	0,8990	1,86	0,9371	2,16	0,9692	2,60	0,9907	3,40	0,9993
1,65	0,9011	1,87	0,9385	2,18	0,9707	2,62	0,9912	3,60	0,99968
1,66	0,9031	1,88	0,9399	2,20	0,9722	2,64	0,9917	3,80	0,99986
1,67	0,9051	1,89	0,9412	2,22	0,9736	2,66	0,9922	4,00	0,999938
1,68	0,9070	1,90	0,9426	0,24	0,9749	2,68	0,9926	4,50	0,999993
1,69	0,9090	1,91	0,9439	2,26	0,9762	2,70	0,9931	5,00	0,99999
1,70	0,9109	1,92	0,9451	2,28	0,9774	2,72	0,9935		
1,71	0,9127	1,93	0,9464	2,30	0,9786	2,74	0,9939		

Scanned with OKEN Scanner

z \ n	2	3	4	5	6	7	8	9	10	11
0,0	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
0,1	532	535	537	537	538	538	538	539	539	539
0,2	563	570	573	574	575	576	576	577	577	577
0,3	593	604	608	610	612	613	614	614	614	615
0,4	621	636	642	645	647	648	650	650	651	651
0,5	648	667	674	678	681	683	684	685	686	686
0,6	672	695	705	710	713	715	716	717	718	719
0,7	694	722	733	739	742	745	747	748	749	750
0,8	715	746	759	766	770	773	775	777	778	779
0,9	733	768	783	790	795	799	801	803	804	805
1,0	0,750	789	804	813	818	822	825	827	828	830
1,1	765	807	824	834	839	843	846	848	850	851
1,2	779	824	842	852	858	862	865	868	870	871
1,3	791	838	858	868	875	879	883	885	887	889
1,4	803	852	872	883	890	894	898	900	902	904
1,5	813	864	865	986	903	908	911	914	916	918
1,6	822	875	896	908	915	920	923	926	928	930
1,7	831	884	906	918	925	930	934	936	938	940
1,8	839	893	915	927	934	939	943	945	947	949
1,9	846	901	923	934	944	947	950	953	955	957
2,0	0,852	908	930	942	949	954	957	960	962	963
2,2	864	921	942	954	960	965	968	970	972	974
2,4	874	931	952	963	969	973	976	978	980	981
2,6	883	938	960	970	976	980	982	984	986	987
2,8	981	946	966	976	981	984	987	988	990	991
3,0	988	952	971	980	985	988	990	992	992	993
3,2	904	957	975	984	988	991	992	994	995	995
3,4	909	962	979	986	990	993	994	995	996	997
3,6	914	965	982	989	992	994	996	996	997	998
3,8	918	969	984	990	994	996	997	997	998	998
4,0	922	971	986	992	995	996	997	998	998	999
4,2	926	974	988	993	996	997	998	998	999	999
4,4	929	976	989	994	996	998	998	999	999	999
4,6	932	978	990	995	997	998	999	999	999	1,000
4,8	935	980	991	996	998	998	999	999	1,000	
5,0	937	981	992	996	998	999	999	1,000		
5,2	940	982	993	997	995	999	999			
5,4	942	984	994	997	998	999	1,000			
5,6	944	985	994	998	999	999				
5,8	946	986	995	998	999	999				
6,0	947	987	995	998	999	1,000				

Scanned with OKEN Scanner

redus (n 40) pentru repartiția Student

z	12	13	14	15	16	17	18	19	20	∞
0,0	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,50000
0,1	539	539	539	539	539	539	539	539	539	53983
0,2	577	578	578	578	578	578	578	578	578	57926
0,3	615	615	616	616	616	616	616	616	616	61791
0,4	652	652	652	652	653	653	653	653	653	65542
0,5	686	687	687	688	688	688	688	688	689	69146
0,6	720	720	721	721	721	722	722	722	722	72575
0,7	751	751	752	752	753	753	753	754	754	75804
0,8	780	780	781	782	782	782	783	783	783	78814
0,9	806	807	808	808	809	809	810	810	810	81894
1,0	831	832	832	833	833	834	834	835	835	84134
1,1	853	854	854	855	856	856	857	857	858	86433
1,2	872	873	874	875	876	876	877	877	878	88493
1,3	890	891	892	893	893	894	894	895	895	90320
1,4	906	907	908	908	909	910	910	911	911	91924
1,5	919	920	921	922	923	924	924	924	925	93319
1,6	931	932	933	934	935	935	936	936	937	94520
1,7	941	943	944	945	945	946	946	947	947	95543
1,8	950	952	952	953	935	955	955	956	956	96407
1,9	958	959	960	961	962	962	963	963	964	97128
2,0	965	967	967	967	968	969	969	970	970	97725
2,2	975	976	977	977	978	979	979	979	980	98610
2,4	982	983	984	985	985	986	986	986	987	99180
2,6	988	988	989	990	990	990	991	991	991	99534
2,8	991	992	992	993	993	994	994	994	994	99744
3,0	994	994	995	995	996	996	996	996	996	99865
3,2	996	996	996	997	997	997	997	998	998	99931
3,4	997	997	998	998	998	998	998	998	998	99966
3,6	998	998	998	999	999	999	999	999	999	99984
3,8	998	999	999	999	999	999	999	999	999	99993
4,0	999	999	999	999	999	1,000	1,000	1,000	1,000	99997
4,2	999	999	1,000	1,000	1,000					99999
4,4	1,000	1,000								99999

Scanned with OKEN Scanner

Valorile raportului dispersiilor F corespunzătoare diferitelor nivele de semnificație

A. Nivel de semnificație 5%.

v_2 \ v_1	1	2	3	4	5	6	8	12	24	00
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,58	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Scanned with OKEN Scanner

Valorile raportului dispersiilor F corespunzătoare diferitelor nivele de semnificaţie

B. Nivel de semnificaţie 1%.

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98,49	99,00	99,17	99,25	99,30	99,33	99,36	99,42	99,46	99,50
3	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,60	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	13,93	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,27	9,89	9,47	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,28	4,86
9	10.56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91
11	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,02	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,59	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,43	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,08	2,65
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,00	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	2,92	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36
22	7,94	5,72	4,82	4,31	3,99	3,76	3,45	3,12	2,75	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,62	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,47	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,29	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60
120	6,85	4,79	3,96	3,48	3,17	2,96	2,66	2,34	1,95	1,38
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,79	1,00

Scanned with OKEN Scanner

Valorile raportului dispersiilor F corespunzătoare diferitelor nivele de semnificație

C. Nivel de semnificație 0,1 %.

v_2 \ v_1	1	2	3	4	5	6	8	12	24	∞
1	405284	500000	540379	562500	576405	585937	598144	610667	623497	636619
2	998,5	999,0	999,2	999,2	999,3	999,3	999,4	999,4	999,5	999,5
3	167,5	148,5	141,1	137,1	134,6	132,8	130,6	128,3	125,9	123,5
4	74,14	61,25	56,18	53,44	51,71	50,53	49,00	47,41	45,77	44,05
5	47,04	36,61	33,20	31,09	29,75	28,84	27,64	26,42	25,14	23,78
6	35,51	27,00	23,70	21,90	20,81	20,03	19,03	17,99	16,89	15,75
7	29,22	21,69	18,77	17,19	16,21	15,52	14,63	13,71	12,73	11,69
8	25,42	18,49	15,83	14,39	13,49	12,86	12,04	11,19	10,30	9,34
9	22,86	16,39	13,90	12,56	11,71	11,13	10,37	9,57	8,72	7,81
10	21,04	14,91	12,55	11,28	10,48	9,92	9,20	8,45	7,64	6,76
11	19,69	13,81	11,56	10,35	9,58	9,05	8,35	7,63	6,85	6,00
12	18,64	12,97	10,80	9,63	8,89	8,38	7,71	7,00	6,25	5,42
13	17,81	12,31	10,21	9,07	8,35	7,86	7,21	6,52	5,78	4,97
14	17,14	11,78	9,73	8,62	7,92	7,43	6,80	6,13	5,41	4,60
15	16,59	11,34	9,34	8,25	7,57	7,09	6,47	5,81	5,10	4,31
16	16,12	10,97	9,00	7,94	7,27	6,81	6,19	5,55	4,85	4,06
17	15,72	10,66	8,73	7,68	7,02	6,56	5,96	5,32	4,63	3,85
18	15,38	10,39	8,49	7,46	6,81	6,35	5,76	5,13	4,45	3,67
19	15,08	10,16	8,28	7,26	6,61	6,18	5,59	4,97	4,29	3,52
20	14,82	9,95	8,10	7,10	6,46	6,02	5,44	4,82	4,15	3,38
21	14,59	9,77	7,94	6,95	6,32	5,88	5,31	4,70	4,03	3,26
22	14,38	9,61	7,80	6,81	6,19	5,76	5,19	4,58	3,92	3,15
23	14,19	9,47	7,67	6,69	6,08	5,65	5,09	4,48	3,82	3,05
24	14,03	9,34	7,55	6,59	5,98	5,55	4,99	4,39	3,74	2,97
25	13,88	9,22	7,45	6,49	5,88	5,46	4,91	4,31	3,66	2,89
26	13,74	9,12	7,36	6,41	5,80	5,38	4,83	4,24	3,59	2,82
27	13,61	9,02	7,27	6,33	5,73	5,31	4,76	4,17	3,52	2,75
28	13,50	8,93	7,19	6,25	5,66	5,24	4,69	4,11	3,46	2,70
29	13,39	8,85	7,12	6,19	5,59	5,18	4,64	4,05	3,41	2,64
30	13,29	8,77	7,05	6,12	5,53	5,12	4,58	4,00	3,36	2,59
40	12,61	8,25	6,60	5,70	5,13	4,73	4,21	3,64	3,01	2,23
60	11,97	7,76	6,17	5,31	4,76	4,37	3,87	3,31	2,69	1,90
120	11,38	7,31	5,79	4,95	4,42	4,04	3,55	3,02	2,40	1,56
∞	10,83	6,91	5,42	4,62	4,10	3,74	3,27	2,74	2,13	1,00

Întotdeauna v_1 corespunde celei mai mari dispersii.

Scanned with OKEN Scanner

BIBLIOGRAFIE

K. Marx — *Capitalul*, vol. I, E.S.P.L.P., București, 1957.

K. Marx — *Capitalul*, vol. II, E.S.P.L.P., București, 1958.

K. Marx — *Capitalul*, vol. III, E.S.P.L.P., București, 1955, partea a III-a.

V.I. Lenin — *Opere complete*, vol. 3, Editura politică, București, 1961.

V.I. Lenin — *Opere*, vol. 22, E.S.P.L.P., București, 1963.

V.I. Lenin — *Opere complete*, vol. 26, Editura politică, București, 1964.

* * * — *Programul Partidului Comunist Român de făurire a societății socialiste multilateral dezvoltate și înaintare a României spre comunism*, Editura politică, București, 1975.

* * * — *Hotărîrea Comitetului Central al Partidului Comunist Român cu privire la perfecționarea sistemului informațional economico-social, introducerea sistemelor de conducere cu mijloace de prelucrare automată a datelor și dotarea economiei naționale cu tehnică de calcul în perioada 1971—1980*, în „Scînteia", 22.IV.1972.

* * * — *Hotărîrea Comitetului Central al Partidului Comunist Român cu privire la perfecționarea conducerii și planificării economico-financiare*, Editura politică, București, 1978.

Nicolae Ceaușescu — *Raport la cel de-al XI-lea Congres al Partidului Comunist Român*, Editura politică, București, 1974.

Nicolae Ceaușescu — *Raport la Conferința Națională a Partidului Comunist Român*, 19—21 iulie 1972, Editura politică, București, 1972.

Nicolae Ceaușescu — *Expunere la deschiderea colocviului privind problemele științei conducerii societății*, Editura politică, București, 1972.

Nicolae Ceaușescu — *Raport la Conferința Națională a Partidului Comunist Român*, 7—9 decembrie 1977, Editura politică, București, 1977.

* * *

Al. Bărbat — *Cu privire la obiectul statisticii*. În „Revista de filozofie" nr. 11, București, 1968.

Al. Bărbat — *Statistica în sistemul științelor sociale*. În „Revista de filozofie" nr. 5, București. 1970.

Al. Bărbat — *Teoria statisticii sociale*, Editura didactică și pedagogică, București, 1972.

M. Biji și colab. — *Curs de Statistică teoretică*, Editura didactică și pedagogică, București, 1963.

M. Biji — *Ion Ionescu de la Brad — un statistician progresist român*, În „Revista de statistică" nr. 5, 1955.

M. Biji, I. Stoichiță — *Metoda selectivă în cercetarea statistică*, Editura științifică, București, 1956.

E. Biji, Simina Popescu Negură — *Metode de calcul al corelațiilor în producția agricolă*, Editura Ceres, București, 1971.

E. Biji, A. Băjan, M. Bădiță, E. Lilea — *Culegere de probleme de teoria statisticii*, Academia de studii economice, București, 1977.

E. Biji, M. Bădiță — *Analiza legăturilor dintre fenomenele social-economice pe plan teritorial, prin metode neparametrice*. În „Revista de statistică", nr. 10, 1972.

E. Biji, M. Bădiță — *Folosirea seriilor cronologice în calcule de perspectivă*. În „Revista de statistică" nr. 4/1973.

G. Gallot — *Cours de statistique descriptive*, Ed. Dunod, Paris, 1965.

T. Cristureanu, I. Ivănescu, I. Negură, G. Stoianovici — *Sistemul informațional și calculul statistico-economic în comerțul exterior*, Editura științifică, București, 1974.

Scanned with OKEN Scanner

F.E. Croxton, D.J. Cowden, S. Klein — *Applied General Statistics*, ed. a 3-a, Prentice Hall, New York, 1967 (traducere parțială în revista „Probleme de statistică social-economică").

I. Désable — *Théorie et practique des sondage*, Ed. Dunod, Paris, 1966.

G. Dunon — *De la théorie à la pratique des indices statistique*, Ed. Gauthier-Villard, Paris, 1955.

A.I. Ejov — *Ajustarea și calcularea șirurilor de distribuție*, D.C.S., București 1963 (traducere din limba rusă).

H. Guitton — *Statistique*, Ed. Dalloz, Paris, 1967.

D. Hașigan, I. Marinescu — *Grafice și elemente de calcul grafic. Metode și aplicații*, Editura științifică, București, 1968.

H.I. Lange — *Extrapolarea șirului dinamice economice după o funcție de ajustare parabolică*, Buletin documentar statistic nr. 2/1970.

I. Marinescu și colectiv — *Elemente de statistică matematică și aplicațiile ei*, Editura științifică, București, 1966.

I. Marinescu — *Despre comparabilitatea șirului dinamice cu variații sezoniere.* În „Revista de Statistică" nr. 9/1971.

I. Marinescu — *Probleme ale regresiei liniare simple.* În „Revista de statistică" nr. 2/1972.

I. Marinescu — *Despre caracterizarea intensității dependențelor statistice din economie.* În „Revista de statistică" nr. 9/1972.

M. Mănescu și colectiv — *Calcul economic*, cap. I și II, Editura Academiei R.S.R., București, 1964.

V. Micev — *Analiza dispersională în cercetarea corelației producției.* În „Buletin documentar statistic" nr. 9/1972.

Gh. Mihoc și V. Urseanu — *Matematici aplicate în statistică*, Editura Academiei R.S.R., București, 1962.

Gh. Mihoc și V. Urseanu — *Sondaj simplu din populații cu o singură caracteristică cantitativă.* În „Revista de statistică" nr. 4, 5, 6/1972.

F.C. Mills — *Metode statistice*, ed. a 3-a, Columbia University, New York, 1955 (traducere Direcția Centrală de Statistică, București).

C. Moineagu — *Modelarea corelațiilor în economie*, Editura științifică, București, 1974.

C. Moineagu, I. Negură, V. Urseanu — *Statistica*, Editura științifică și enciclopedică, București, 1977.

C.A. Moser — *Metode de anchetă în investigarea fenomenelor sociale*, Editura științifică, București, 1967.

A. Piatier — *Statistique descriptive et initiation a l'analise.* Presses Universitaires de France, Paris, 1962.

N. Rancu, L. Tövissi — *Statistică matematică cu aplicații în producție.* Editura Academiei R.S.R., București, 1963.

G. Milton Smith — *Ghid simplificat de statistică pentru psihologie și pedagogie*, Editura didactică și pedagogică, București, 1971.

L. Tövissi — *Probleme teoretice ale calculului indicilor teritoriali.* În „Studii și cercetări economice", vol. II, București, 1964.

D. Ticulescu, E. Biji — *Statistică aplicată în economia întreprinderilor agricole socialiste*, Editura Ceres, București, 1975.

G. Undy Yule, M.G. Kendall — *Introducere în teoria statisticii*, ed. XIX, Editura științifică, București, 1969.

* * * — *Allgemeine Statistik*, Verlag die Wirtschaft, Berlin, 1965.

* * * — *Metode statistice de cercetare a corelațiilor în economie*, Editura Academiei R.S.R., București. 1967.

* * * — Colecția „Teorie și metodă în științele sociale", vol. I, vol. II, Editura politică, București, 1965—1966.

Scanned with OKEN Scanner

ERATA

Pag.	Rîndul	În loc de:	Se va citi:
123	13 de sus (relaţia 6.39)	$\bar{x}_g = \sqrt{\prod_{i=1}^{n} x_i}$	$\bar{x}_g = \sqrt[n]{\prod_{i=1}^{n} x_i}$
145	7 de sus (relaţia 6.94)	$\left(\frac{\sum (x_i'' - \bar{x}'')^2}{n}\right)\frac{1}{K^2} = \frac{\sigma_x^2}{K^2}$	$\frac{\sum (x_i'' - \bar{x}'')^2}{n} = \frac{\sigma_x^2}{K^2}$
152	ultimul (ultima fracţie)	$\frac{\sum_{i=1}^{r} (\bar{x}_i - \bar{x}_0) f_i n_i}{\sum_{i=1}^{r} n_i}$	$\frac{\sum_{i=1}^{r} (\bar{x}_i - \bar{x}_0)^2 n_i}{\sum_{i=1}^{r} n_i}$
252	9 de sus	$\sigma_{y/x}^y$	$\sigma_{y/x}^2$
300	11 de sus (fracţiile 2 şi 3)	$\frac{y_i}{y_{i-1}} = \frac{y_{i-1}}{y_{i-1}}$	$\frac{y_i}{y_{i-1}} - \frac{y_{i-1}}{y_{i-1}}$
307	19 de sus (relaţia 10.25)	$\sqrt{\prod_{i=1}^{n} I_{i/i-1}} = \sqrt[n]{I_{i/0}} = \sqrt{\frac{y_n}{y_0}}$	$\sqrt[n]{\prod_{i=1}^{n} I_{i/i-1}} = \sqrt[n]{I_{n/0}} = \sqrt[n]{\frac{ny}{y_0}}$
356	12 de sus	143,2 ori	143,2 %
358	tabelul 10.20 (rîndul 2)	773	23 773
403	tabelul 11.5 (rîndul 1, col. 4)	3 700	3 780
404	5 de jos (relaţia 11.107)	$\frac{1}{I_{1/0}}$	$\frac{1}{I_{1/0}^t}$

Scanned with OKEN Scanner

Plan editură 5843. Tiraj 5400+90 ex. leg. 1/1
Coli de tipar 27. Bun de tipar 16.II.79

C. *1860* — I. P. INFORMAȚIA
str. Brezoianu nr. *23—25*, București
Republica Socialistă România

Scanned with OKEN Scanner